REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
Nov. 14, 1996

3. REPORT TYPE AND DATES COVERED
Final Technical/ 15-8-95/14-8-96

4. TITLE AND SUBTITLE
Target Acquisition in Complex Scenes
Part A: Search and Conspicuity Models

5. FUNDING NUMBERS
G
F49620-95-1-0495

6. AUTHOR(S)
A. Toet

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TNO Human Factors Research Institute
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
EOARD/ALB
223/231 Old Marylebone Road
London NW1 5TH
United Kingdom

10. SPONSORING /MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release
Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A visual search and detection experiment is performed on a set of complex natural images with military vehicles as targets. The area under the resulting cumulative detection probability curve of each target is adopted as a characteristic measure for its visual distinctness. The visual distinctness rank order induced by this measure is adopted as the reference rank order. This study investigates the capability of several digital target distinctness metrics and the psychophysically determined target visual lobe (i.e. the minimal distance between target and eye-fixation at which the target is no longer distinguishable from its surroundings) to reproduce the abovementioned reference rank order.

The visual lobe indeed appears a useful predictor of human performance in a visual search and detection task. Models of the early human visual system, a normalised root-mean square metric, and the edge distance metric introduced in this report, all seem to induce a visual distinctness rank ordering that agrees with human visual perception. Metrics based (1) on first order statistics of the graylevel histogram, (2) on the intersection of (oriented) graylevel histograms of target and background, and (3) on a combination of area and edge contrast, all correlate poorly with human observer performance. The CAMAELEON model (based on histogram intersection) is also highly sensitivity to variations in the definition (size and shape) of the target and background masks.

The Perceptual Distortion model induces a visual target distinctness rank ordering identical to the one resulting from human observer performance, and therefore shows the best overall performance of all models and metrics tested in this study.

14. SUBJECT TERMS
Conspicuity; saliency; search; target distinctness; visual lobe; earley vision models.

15. NUMBER OF PAGES
73

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
unclassified	unclassified	unclassief	UL

Target Acquisition in Complex Scenes

Part A: Search and Conspicuity Models

FINAL TECHNICAL REPORT

by

Dr. A. Toet

November 14, 1996

United States Air Force

EUROPEAN OFFICE OF AEROSPACE RESEARCH AND DEVELOPMENT

London England

GRANT NUMBER F49620-95-1-0495

TNO Human Factors Research Institute, Soesterberg, The Netherlands

CONTENTS

SUMMARY

The present pilot study investigates the capability of (i) several digital image metrics that have been reported in the literature, (ii) the digital edge distance metric defined in this report, and (iii) the psychophysically determined target visual lobe (i.e. the minimal distance between target and eye-fixation at which the target is no longer distinguishable from its surroundings), to predict human visual search performance for a small number of realistic and military relevent complex scenes.

A visual search and detection experiment is performed on a set of complex naturalistic images with military vehicles as targets. The area under the resulting cumulative detection probability curve of each target is adopted as a characteristic measure for its visual distinctness. The visual distinctness rank order induced by this measure is adopted as the reference rank order. The targets are also rank-ordered with respect to their visual distinctness as determined by the visual lobe and the different computational metrics. The rank orders produced by the psychophysical and computational target distinctness measures are compared with the reference rank order. The following conclusions can be drawn from the results of this exercise.

The visual lobe indeed appears a useful predictor of human performance in a visual search and detection task. Models of the early human visual system (Cortex, TARDEC, Perceptual Distortion), a normalised root-mean square metric, and the edge distance metric introduced in this report, all seem to induce a visual distinctness rank ordering that *agrees* with human visual perception. Metrics based (a) on first order statistics of the graylevel histogram (Doyle), (b) on the intersection of (oriented) graylevel histograms of target and background, and (c) on a combination of area and edge contrast, all *correlate poorly* with human observer performance. The CAMAELEON model (based on histogram intersection) is also highly sensitivity to variations in the definition (size and shape) of the target and background masks.

The Perceptual Distortion model induces a visual target distinctness rank ordering identical to the one resulting from human observer performance, and therefore shows the best overall performance of all models and metrics tested in this study.

1 INTRODUCTION

Aim

The present pilot study is performed to investigate which visual target signature characteristics can be used as predictors of human visual search performance in realistic and military relevant complex scenes.

Approach

A visual search and detection experiment is performed on a set of complex natural images with military vehicles as targets. The area under the resulting cumulative detection probability curve of a target is adopted as a characteristic measure for the visual distinctness of the target. The visual distinctness rank order induced by this measure is adopted as the reference rank order. One newly developed psychophysical procedure and several digital image analysis algorithms are applied to determine visual target distinctness for a small subset of the images used in the visual search experiment[1]. The targets are rank-ordered with respect to their visual distinctness as determined with these different approaches. Finally, for each target distinctness measure investigated, the correlation is determined between the rank order induced by each measure and the abovementioned reference rank order.

Background

It is of great military importance to have advance knowledge of human acquisition performance for targets or other relevant objects. Advance knowledge of the time an observer needs to find and detect a target is of interest in preparing real flight scenarios, in modelling mission performance, and in evaluating camouflage effectiveness. However, search performance inherently shows a large variance, and depends strongly on familiarity with the perceived scene. Therefore, in a search experiment each observer should only perform once in each particular scenario. Mean search performance can only be determined by averaging over many observers. In field situations measuring acquisition performance is usually impractical and often too costly or even too dangerous. It is therefore of great practical value to have measures that (i) reliably predict human visual search performance and that (ii) are easy to determine, without large cost or risk. The objective of the present study is to perform a pilot validation of the capability of a newly developed psychophysical measure and several digital image analysis methods to predict human target acquisition performance.

[1] The comparative model study depended on the voluntary cooperation of a number of researchers that were willing to run their models on the selected set of test images. To limit the time requirements the number of test images necessarily had to be small.

Target distinctness and visual search

Target acquisition is a complex process, and many factors involved are not yet fully understood. One thing is evident: the more a target stands out from its background the easier it will be to detect it. It is therefore a priori likely that a contrast measure that captures a target's visual distinctness as perceived by a human observer should correlate with human visual search performance.

The psychophysical method and the computational approaches investigated in this study represent different approaches to the problem of quantifying the *visual distinctness* of a target in its surround.

Factors determining target distinctness

An object can be distinguished or detected visually when its retinal image differs in some way from the image of its surroundings. It is known that the human visual system utilizes differences in size, shape, luminance, colour, texture, binocular disparity and motion to achieve this figure-ground segregation, and that any of these factors is sufficient on its own.

The visual system predominantly analyses local *feature differences.* Discrimination performance depends on the amount of the feature differences. Detection, classification, recognition and identification are merely progressive levels of discrimination. Each of these levels of discrimination corresponds to a higher order feature difference between the target and its local surround. Local *feature contrast* should exceed the overall variation in a pattern to allow a target to stand out from its background and be detected (Alkhateeb et al., 1990a; Cole & Jenkins, 1984; Jenkins & Cole, 1982; Nothdurft, 1991c, 1992, 1993a,b,c). Pronounced feature differences fail to produce visual pattern segregation when the variations are made continuous (Nothdurft, 1985b, 1990, 1992, 1993a,b). Similarly, the visibility of a motion defined target depends on the local motion contrast of the target elements (i.e. on the magnitude of the motion gradient along the borders of the target support; Nothdurft, 1993b; Sachtler & Zaidi, 1995). Early feature segregation is probably mediated by contrast sensitive striate neurons (Kastner et al., 1997).

The segregation of a region of visual space from its surround is an important part of early (preattentive) vision. One hallmark of this process is that it is parallel, independent of focussed attention (Bergen & Julesz, 1983). The parallel preattentive or bottom-up stage is thought to guide a serial (computationally intensive) attentive or top-down stage. Current models of human visual search and detection assume that the preattentive stage indicates potentially interesting image regions, whereupon the focus of attention is sequentially shifted to each of these regions and the serial stage is deployed to analyze them in detail (Doll et al., 1993; Koch & Ullman, 1985; Olshausen et al., 1993; Rybak et al., 1993; Tsotsos, 1990, 1993, 1994; Wolfe, 1992, 1994a,b; Wolfe & Cave, 1989; Wolfe et al., 1989).

The visual cues that give rise to perceptual segregation can be distinguished into low- and high- level cues. *Low level cues* are for instance differences in luminance and color.

These may already be encoded at a retinal level. *Higher order cues* are for instance local differences in texture, motion or stereo disparity. These require further analysis before local differences can be detected. Good segregation has for instance been found for texture patterns with line arrays at different orientations (Beck, 1966a, 1972, 1982; Julesz, 1975, 1984; Olson & Attneave, 1970; Nothdurft, 1985b), for dot patterns with dots either moving in different directions (Nakayama & Tyler, 1981; Golomb et al., 1985; Nakayama et al., 1985; Nothdurft, 1987), or seen at different disparity (Julesz, 1960, 1964, 1971), for targets that differ from the other elements in the pattern in their orientation (Treisman & Gormican, 1988; Sagi & Julesz, 1987; Foster & Ward, 1991; Nothdurft, 1991b, 1992; Beck, 1966a, b, 1972), motion (Nakayama & Silverman, 1986; Dick et al., 1987), disparity (Nakayama & Silverman, 1986), and size (Jenkins & Cole, 1982; Cole & Jenkins, 1984).

Psychophysical distinctness measure

Target distinctness can operationally be defined as the peripheral area around the central fixation point from which specific target information can be extracted in a single glimpse (Engel, 1971, 1974, 1977). This area is sometimes referred to as the *visual lobe* or *conspicuity area.* The size and shape of the conspicuity area have been measured for a range of static targets in static scenes (Bloomfield, 1972; Bowler, 1990; Cole & Jenkins, 1984; Engel, 1971, 1974, 1977; Jenkins & Cole, 1982). It is found that the conspicuity area is small if the target is embedded in a complex background (a surround with high feature variability) or if the target is surrounded by irregularly positioned nontargets of high similarity (a surround with high spatial variability). The conspicuity area is large if the target stands out clearly from a homogeneous background. The abovementioned definition of target distinctness therefore leads to results that agree with the intuitive notion of target detectability. However, the psychophysical procedure that was originally used to measure the visual lobe of a target is rather intricate and time consuming (Engel, 1971, 1974, 1977).

The TNO Human Factors Research Institute recently developed a novel psychophysical procedure to measure the visual distinctness of a target. In this approach, target distinctness is experimentally defined as the minimal distance between target and eye-fixation at which the target is no longer distinguishable from its surroundings (Wertheim, 1989). The measurent procedure is extremely simple and fast (for detailed description see Section 3.2). Pilot experiments have shown a high degree of correlation between mean search time and target distinctness determined with this new procedure (Kooi & Valeton, 1997). This implies that average search times can in principle be predicted from this target distinctness measure.

Computational distinctness measures

Current models of human visual search are based on the outcome of laboratory experiments using simple artificial stimuli, presented under extremely restricted (impoverished) conditions, and in different experimental paradigms. The construction of most models involves a large number of assumptions, extrapolations and educated guesses. Each of the rules that are employed is based on the study of a particular aspect of the human visual search capability. It is not clear to which extent the combination of these facts is a valid characterization of the overall search process in a complex environment. Consequently, most current target acquisition models fail to predict actual observer acquisition performance in military relevant realistic scenarios.

Models of the human visual search and detection capability that predict the detection probability for certain objects as a function of time usually approximate the conspicuity area of a given target by the contrast detection threshold as a function of eccentricity, corresponding to a square or disc with an angular extent equal to that of the target and presented on a homogeneous background (Bowler, 1990; Overington, 1982; Kraiss & Knäeuper, 1982; Waldman et al. , 1991). Consequently, these models cannot predict actual observer performance on the acquisition of complex targets in complex backgrounds (e.g. Bijl, 1996; Bijl & Valeton, 1994).

To extend these models to arbitrary visual inputs, it needs to be known how the local differences along each of its characteristic dimensions (i.e. feature differences) contribute to an object's complex contrast or saliency. Some theoretical and experimental approaches to estimate the conspicuity area (Cole & Jenkins, 1984; Jenkins & Cole, 1982) or the complex contrast (Peli, 1990; Lillesæter, 1993) of (structured) targets on complex backgrounds have been reported.

Overview

The rest of this report is organised as follows.
Section 2 describes the registration procedure and the selection of the images of complex natural scenes that are used to evaluate the psychophysical and computational target distinctness measures.
Section 3 presents two observer experiments that are performed to obtain psychophysical target distinctness measures for the images described in Section 2.
Subsection 3.1 describes the experimental procedure and the results of the experiment that is performed to test observer visual search performance with the abovementioned imagery.
Subsection 3.2 presents the psychophysical procedure that is used to measure the visual distinctness of the targets in the images that were used in the visual search experiment.
Section 4 reviews the computational target distinctness measures and early vision models that are currently available, and introduces a new target distinctness metric based on the density and local spatial layout of the edges in the target surround (Subsection 4.19). The algorithms that are a priori most likely to produce results that correlate with human performance are selected and applied to a small subset of the images that are used in the visual search experiment.
Section 6 presents a general discussion of the results of the present study.
Section 7 summarizes the main findings of this study.
Section 8 suggests a follow-up project to further evaluate the outcome of the present pilot study.

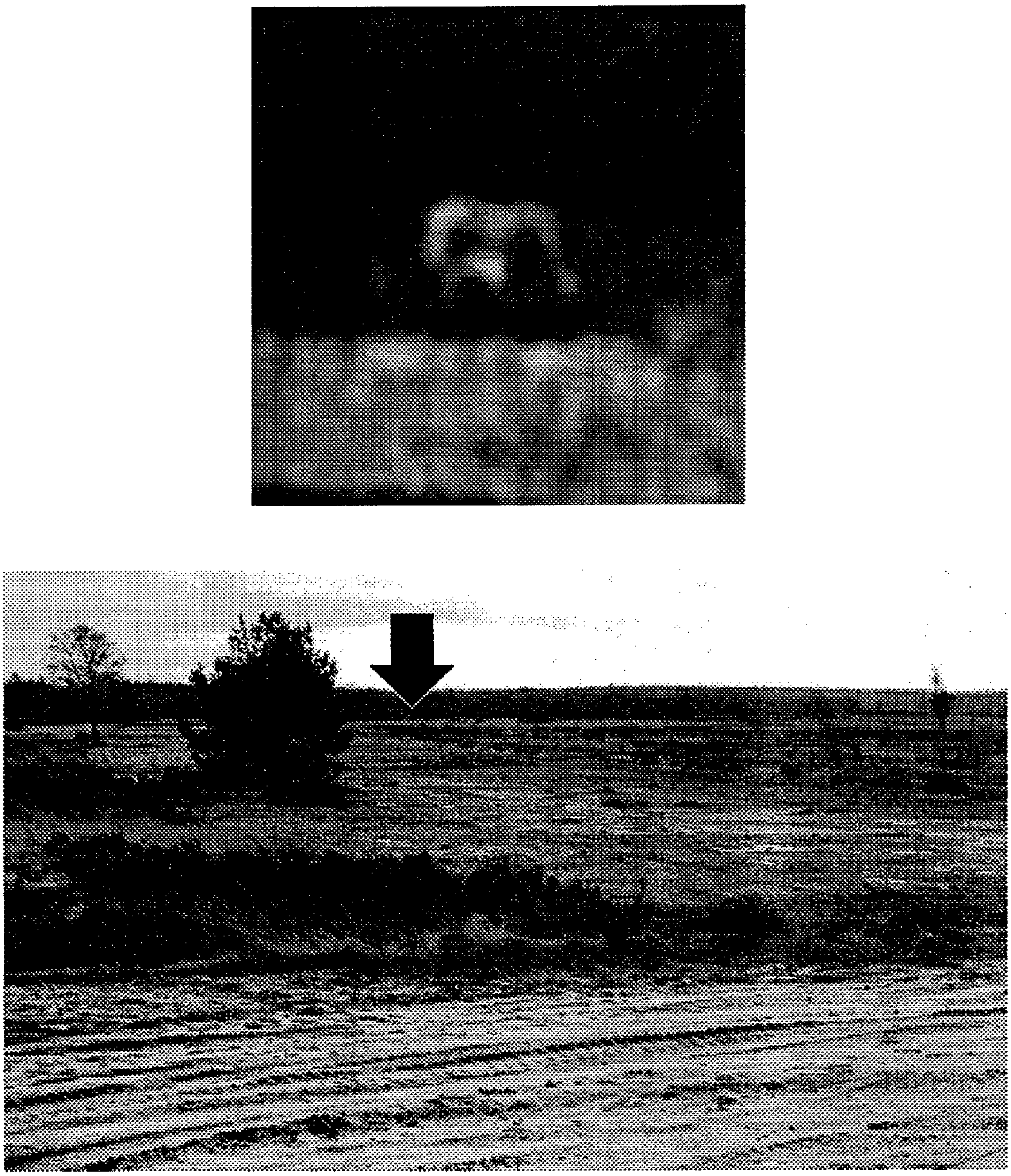


Fig. 1 Complex rural scene with a Land Rover vehicle that serves as a search target. The upper image is an enlargement of the target section in the lower image. The position of the target is indicated by the arrow.

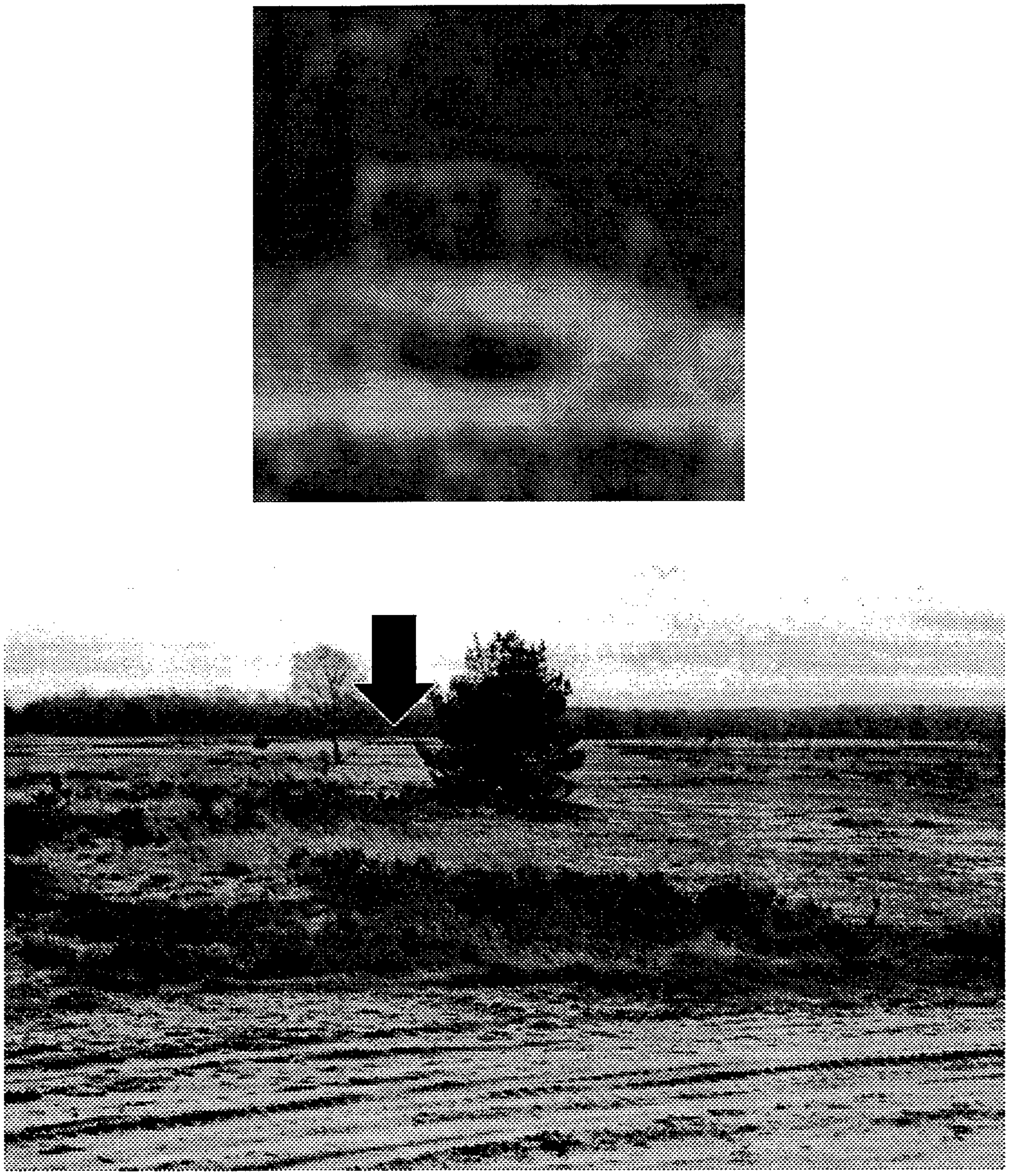


Fig. 2 As Figure 1.

Fig. 3 As Figure 1.

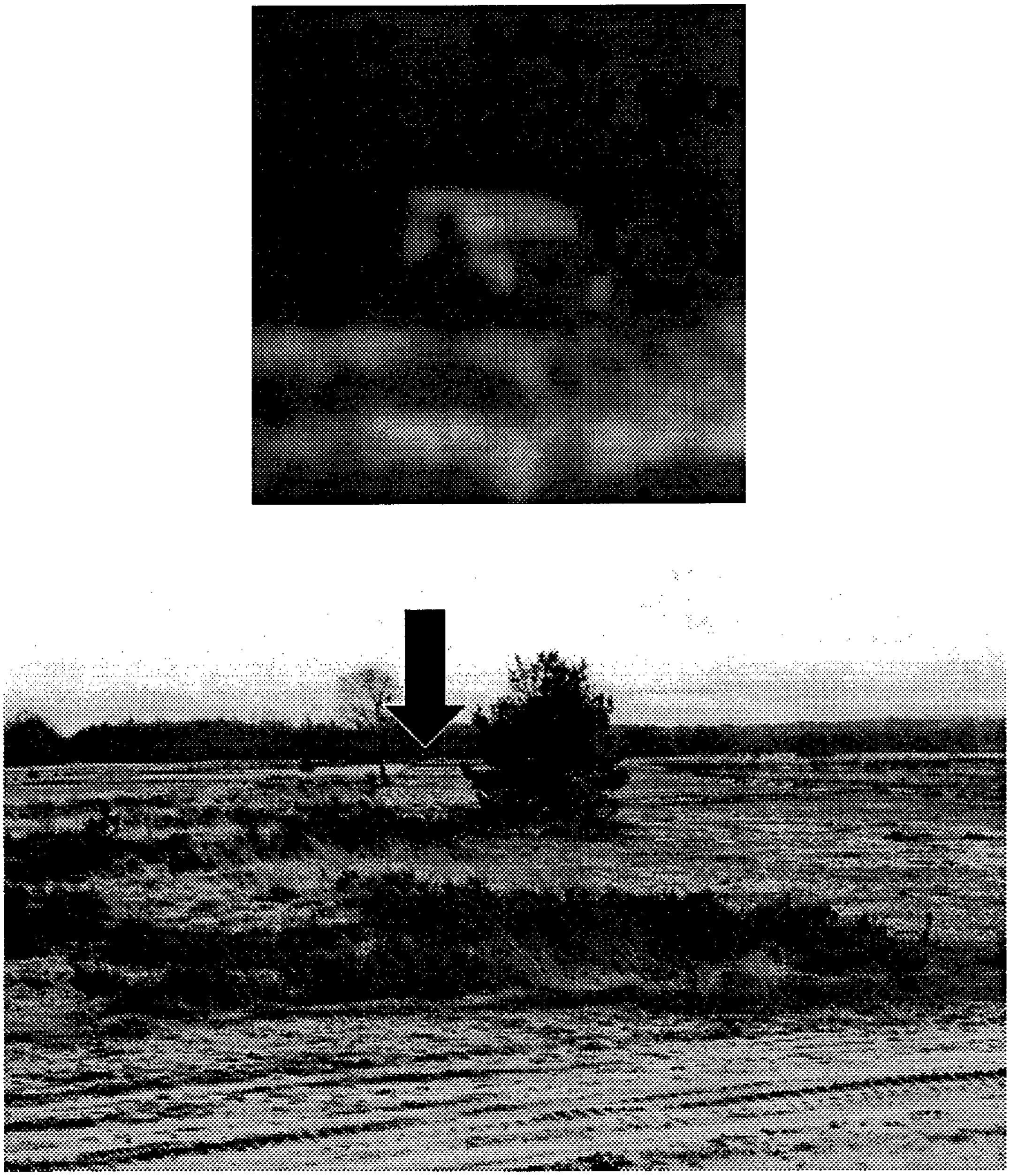


Fig. 4 As Figure 1.

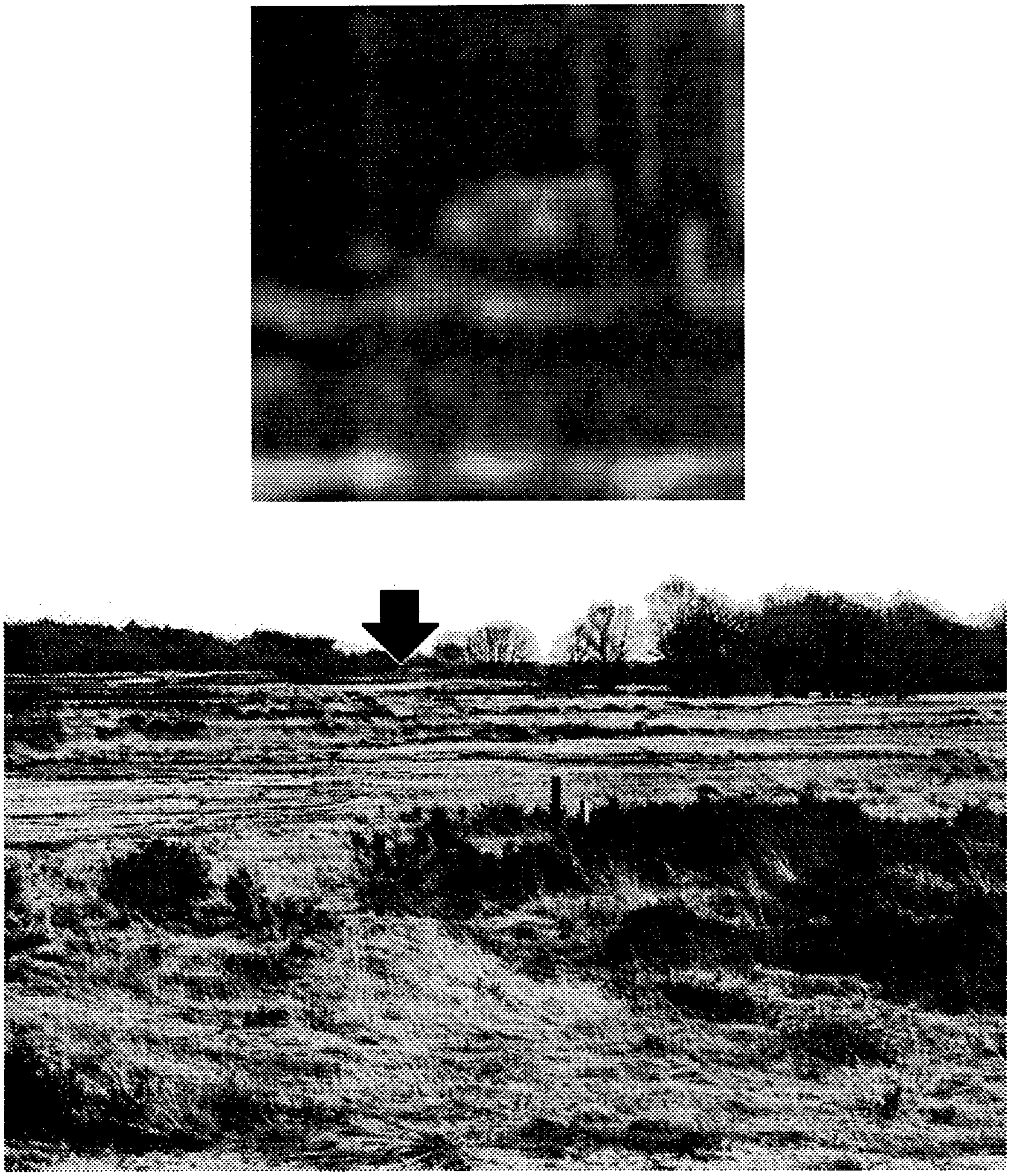


Fig. 5 As Figure 1.

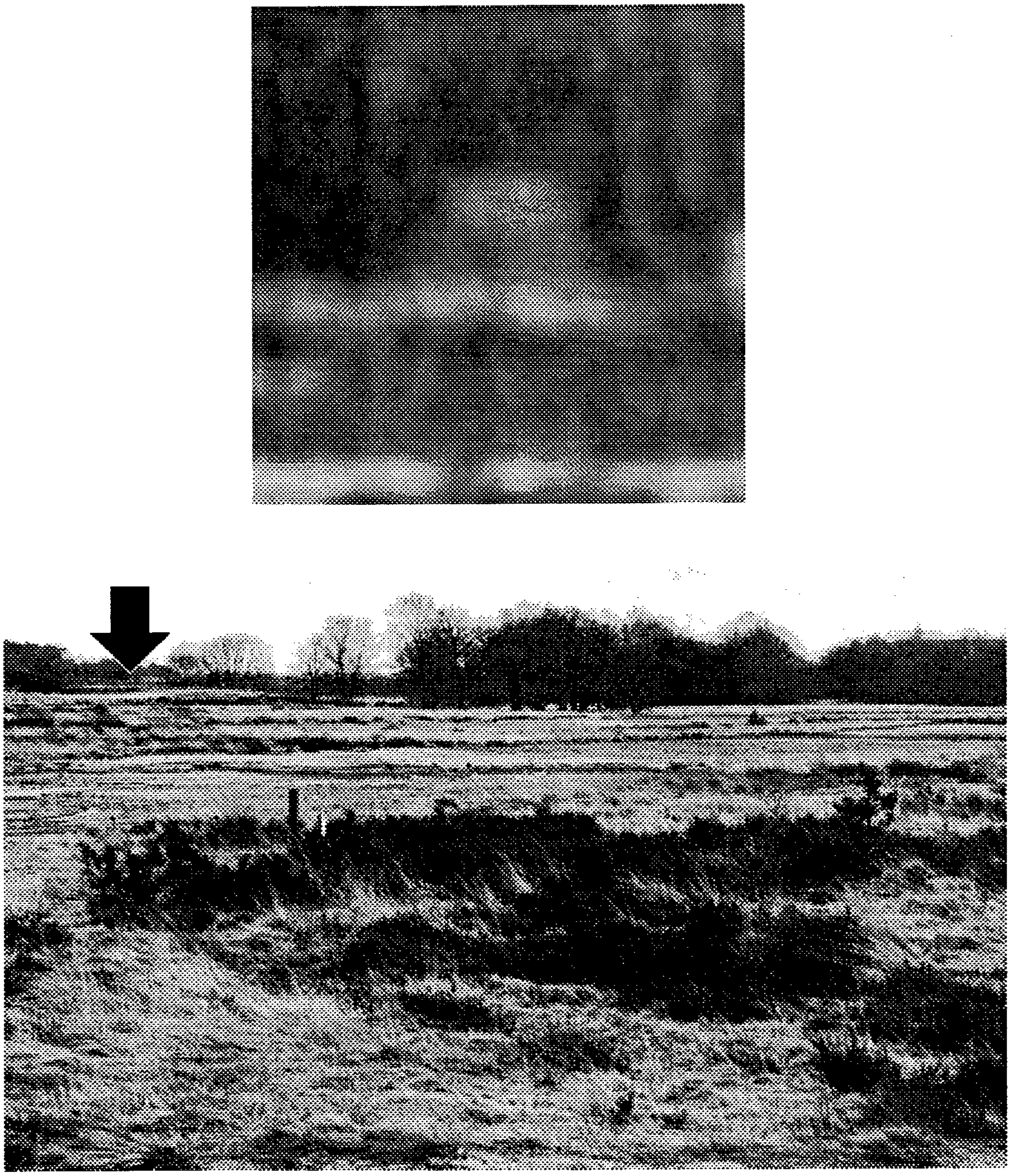


Fig. 6 As Figure 1.

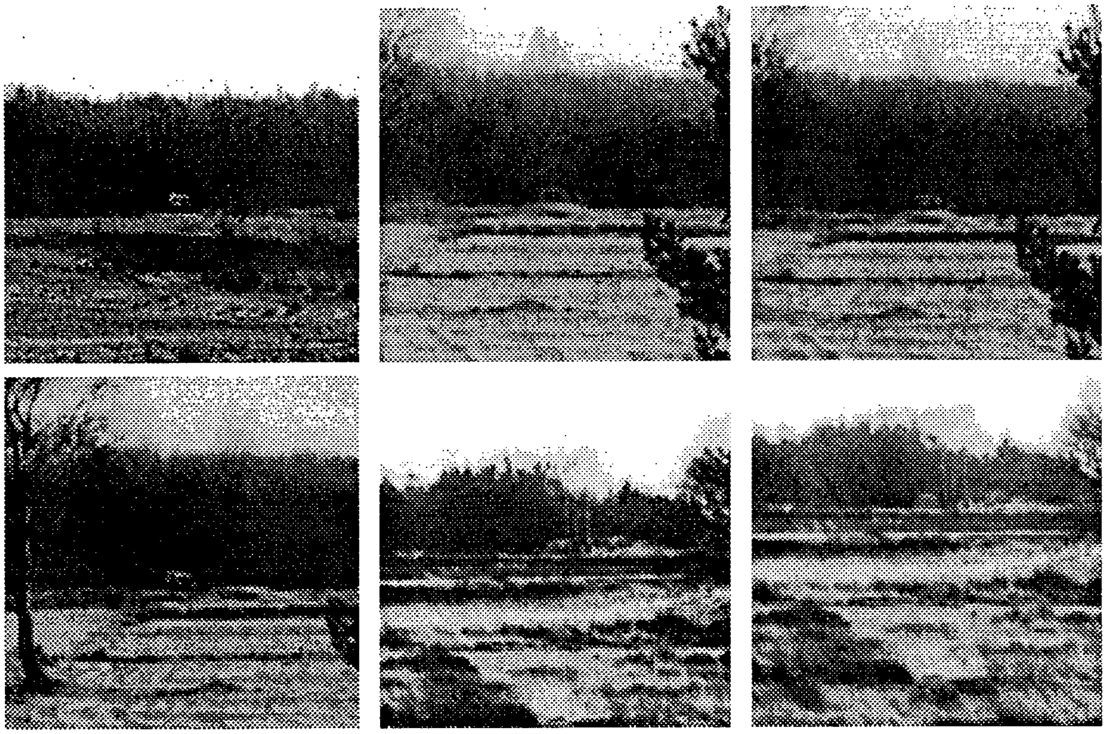


Fig. 7 Target sections (of size 256 × 256 pixels) from Figs. 1–6 . (from upper left to lower right).

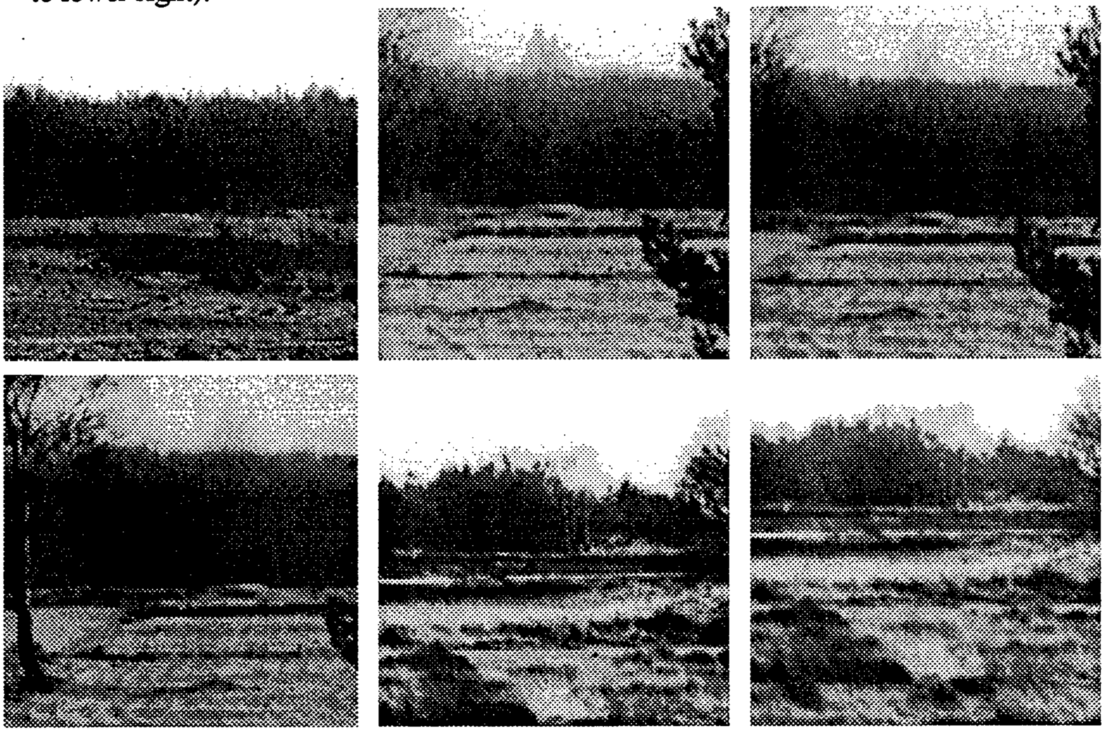


Fig. 8 Target sections (of size 256 × 256 pixels) of empty scenes corresponding to Figs. 1–6 (from upper left to lower right).

Table I Data corresponding to Figs. 1–6.

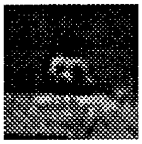
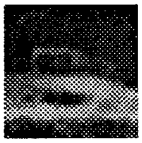
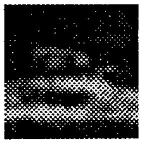
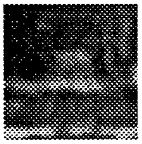
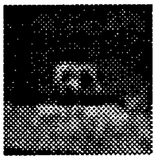
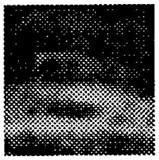
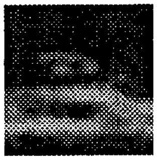
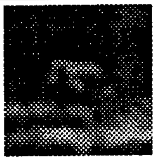
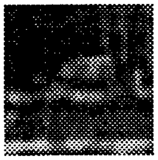
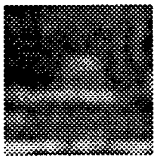
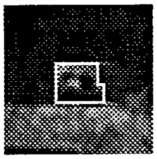
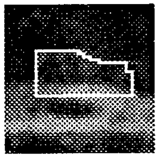
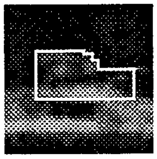
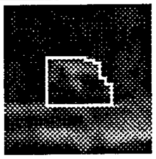
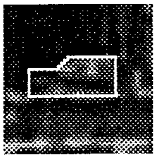
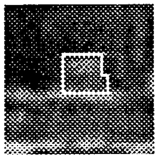
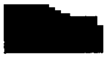
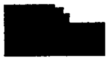
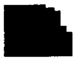
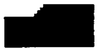
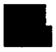
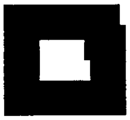
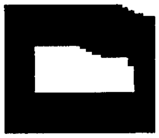
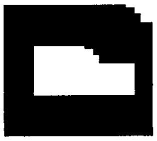
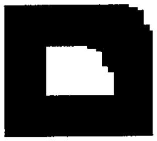
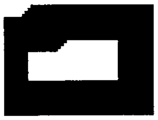
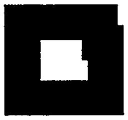
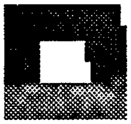
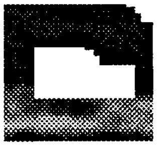
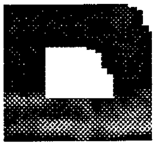
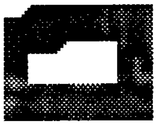
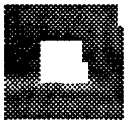
		Scene parameters			Display parameters			Digital parameters					
Fig.	target section	dist.	view	camou-flage	width	L_t	L_b	μ_t	μ_b	μ_p	σ_t	σ_b	σ_p
		(m)			(min)	(cd/m^2)	(cd/m^2)	(–)	(–)	(–)	(–)	(–)	(–)
1		640	front	no	23	4.9	4.3	69.2	77.9	76.5	45.8	39.2	40.5
2		500	side	yes	34	5.5	5.1	95.8	90.4	91.4	33.1	43.0	41.4
3		500	side	no	40	5.9	4.6	84.0	76.5	77.8	30.2	49.1	46.5
4		500	front	no	31	2.5	2.0	73.8	61.8	63.2	31.5	36.7	36.3
5		945	side	no	39	4.1	3.0	96.2	73.6	76.7	25.2	24.0	25.4
6		945	front	yes	29	1.9	1.4	122.9	95.4	99.2	24.0	19.3	22.2

Table II Definition of the target and local background sections of Figs. 1–6.

	Fig. 1	**Fig. 2**	**Fig. 3**	**Fig. 4**	**Fig. 5**	**Fig. 6**
target region						
target outline						
target mask						
background mask						
target area						
background area						

2 IMAGES

All computational experiments in this study are performed on the 6 graylevel (B/W) images shown in Figs. 1–6. These images are selected from a set of 65 gray-scale images that was used to study human visual search and detection performance on realistic complex imagery (see Section 3). Each image represents a Land Rover vehicle in a complex rural background. The exact position in the scene, the viewing distance, and the orientation of each vehicle are known. In Figures 1–6 the location of the target is indicated by arrows. The upper part of Figs. 1–6 shows an enlarged representation of the target area and its immediate surround.

The target was photographed at the Ermelose Heide near Ermelo, The Netherlands. The viewing distance varies between 300 and 1200 m. Images were captured with the target vehicle in different positions relative to the image plane, and both with and without camouflage. As a result the target definition in these images ranges from fairly visible to almost invisible. All images were taken with a standard 50 mm lens, and registered on Agfa Ortho 12 ISO 36 mm B/W slides.

A scanner is used to digitise the slides. The images are quantized to 8 bits. Thus, the radiances for all scene elements were normalized to 255 gray value levels. All measurements and calculations performed on the scenes were made in terms of this gray scale. The size of the digitised image files is 2766 × 1666 pixels. The field-of-view of the 50 mm lens is about 45 degrees. The width of a single pixel is therefore about 0.976 (= 45/2766) min of arc.

The current selection (Figs. 1–6) is made by a subjective inspection and comparison of enlarged versions of the target representation in both large screen-projections of the slides and in zoomed-in CRT projections of the digitized images. In the latter case, digital edge detectors were employed to get an impression of the articulation and the completeness of the target contours (see Fig. 12, page 44). The intent is to create a small sample set with a large variation in visual target distinctness. The results of this subjective selection procedure appear to correlate with the median visual search time that was determined for each particular image from the results of the search and detection experiment (see Section 3.1).

Some computational approaches to the quantification of visual target distinctness need to compare the target scene to exactly the same scene without the target (empty scene). Figure 8 shows the target sections of Figs. 1–6 and the same sections of the corresponding empty scenes. The empty scene is everywhere equal to the target scene, except at the location of the target, where the target is replaced with the local background. This replacement is done by hand, using the rubber stamp tool in Photoshop 3.05. The result is judged by eye and is accepted if the variation in the background over the target support area does not appear to have an appreciable contrast with the natural variation in the local background.

Some computational methods require a precise definition of the target support. Other methods need a definition of the local background (e.g. Hecker, 1992) or the definition of a zone within which background elements are able to interfere with target details (e.g. Tidhar et al., 1994). Table II shows the construction of the target and background mask images for Figs. 1–6. These masks are used in all further processing to restrict the computations to the visible parts of the target support and its local background.

Most algorithms that compute target distinctness from digital imagery only apply to the target support and its immediate surround. Figure 7 shows the 256 × 256 pixels wide target sections of Figs. 1–6. These subimages are centered on the midpoint of the target support (i.e. the window is chosen such that the target is always in the center of the subimage).

3 PSYCHOPHYSICAL TARGET DISTINCTNESS MEASURES

3.1 Search experiment

Search times and cumulative detection probabilities are measured for a military target vehicle (Land Rover) in a complex natural background.

Stimuli

The stimuli are 65 different B/W slides of a Land Rover vehicle in a rural scene (6 of these images are shown in Figs. 1–6). The slides are taken for different positions and orientations of the vehicle in the scene. In some cases camouflage nets are applied to reduce the visibility of the target. Because of the variations in the structure of the local background and the orientation of the target the visibility of the target varies largely throughout the entire stimulus set.

Apparatus

A Kodak Ektapro 7000 caroussel slide projector, equipped with a 90 mm lens, is used to project the slides onto a white screen. A second projector is used to create a bright boundary around the projected scene.

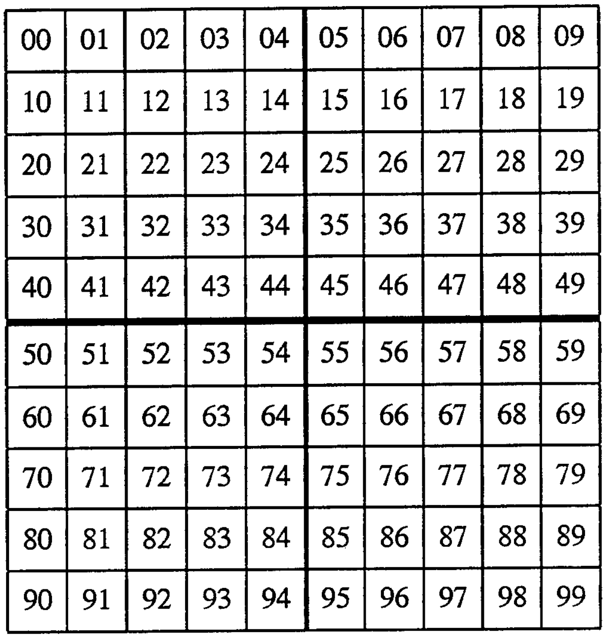
A PC is used to control the order and duration of the presentation of the stimuli and to register the response times and the estimated target locations.

Procedure

A search trial starts with the presentation of a new scene. The subject's taks is to press the space bar of the computer keyboard immediately upon detection of the target. The temporal interval that elapses between the onset of the displayed scene and the moment the subject indicates that the target has been found (by pressing the space bar) is registered and adopted as the search time. The displayed scene disappears immediately after the subject indicates that the target is found and a slide showing a 10 × 10 grid with cells numbered from 0 to 99 is displayed instead. The subject is then requested to indicate the perceived location of the target by entering the number of the grid cell that covers the location at which the target was previously seen.

Subjects

Ten subjects, aged between 37 and 50 years, serve in the experiments reported below. All subjects have (corrected to) normal vision.



00	01	02	03	04	05	06	07	08	09
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	58	59
60	61	62	63	64	65	66	67	68	69
70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89
90	91	92	93	94	95	96	97	98	99

Fig. 9 Image of the grid slide used to indicate the perceived location of the search targets.

Viewing conditions

Viewing is binocular. The experiments are performed in a dimly lit room. The slides are projected onto a wide screen subtending 45 degrees at a viewing distance of 3 m. Since the images were registered with a standard 50 mm lens the scene is represented at true angular size.

Results and discussion

Figure 10 shows the cumulative distribution functions corresponding to the search times measured for Figures 1–6. The overall difference between two of these functions can be measured by subtracting the area beneath their graphs. This operation corresponds to a Kolgomorov-Smirnov (K-S) test. To compare the relative distinctness of the targets in the different scenes the curves are rank-ordered according to the area beneath their graphs. The resulting rank order is listed in the column with the header $\boldsymbol{R}_{\mathrm{Pd}}$ in Table III. This rank order is adopted as the reference standard for the evaluation of (i) the new psychophysical procedure to measure visual target distinctness (see next subsection) and (ii) a selected number of computational target distinctness measures (as described in Section 4; for the evaluation of the different metrics see Section 5).

Targets that give rise to closely spaced cumulative detection curves that cross each other have similar visual distinctness. Figure 10 shows that the test images shown in Figures 1–6 are clustered into three sets of targets with comparable visual distinctness: {1,3,4}, {2}, and {5,6}. Rank order permutations of elements of the same cluster are therefore not very significant. Rank order permutations of elements of different clusters are significant.

The remaining results will be discussed in Section 5, together with the results of the selected computational target distinctness measures.

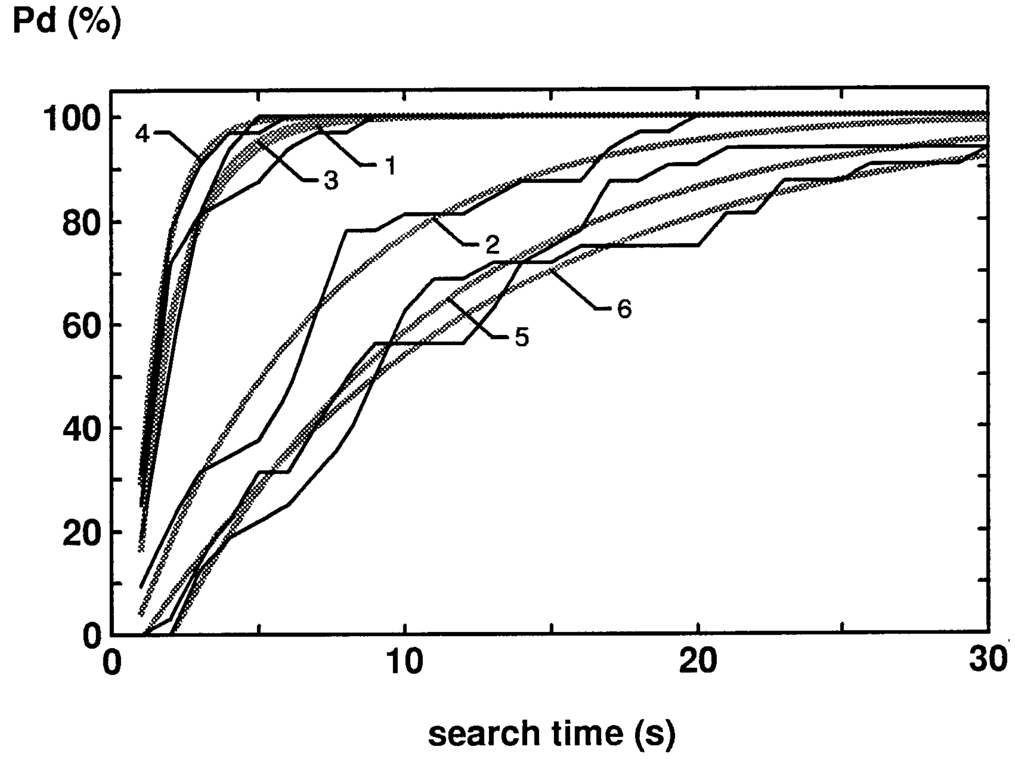


Fig. 10 Cumulative detection probability Pd (%) as a function of search time (s) for the targets in Figures 1–6. The thin black lines represent the experimental data. The thick gray lines represent non-linear least-squares fits of $Pd(t) = 1 - e^{-(t-t_0)/\tau}$ (see Eq. 41) to the experimental data. The curve fits are labelled with the numbers of the corresponding Figures (1–6).

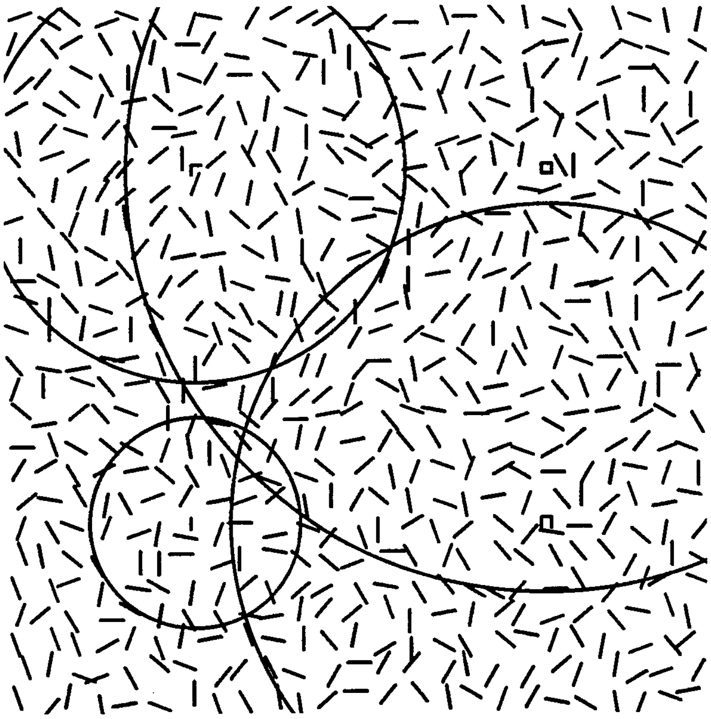


Fig. 11 Illustration of the concept of visual lobe. The circles represent the extent of the visual lobe of the target in the center, and correspond to the maximal separation between fixation and the target at which the target can still be distinguished from its local background. The □ target (upper right) has the largest lobe, followed by the ⊓ shaped target (lower right), and by the ∧-shaped target. The small line segment (lower left) has the smallest visual lobe.

3.2 Lobe experiment

Background

Target conspicuity is operationally defined as the maximum distance between target and foveation and measured in the fronto-parallel plane through the target[2] at which the target still appears distinct from its surround. This conspicuity measure has been shown to be (a) independent of viewing distance, (b) consistent among observers, and (c) meaningful in the sense that it is related to search time (Kooi & Valeton, 1997). The conspicuity distance is easy and quick to measure and can be used on familiar scenes.

Figure 11 illustrates the concept of conspicuity distance. This Figure shows four different targets (respectively □, ⊓, and ∧-shaped, and a small line segment) embedded in a field of irregularly distributed line segments. The circles represent the extent of the visual lobe of

[2] The plane that is parallel to the image plane and at the same distance from the observer as the target.

the target in their center, and correspond to the maximal separation between fixation and the target at which the target can still be distinguished from its local background. The □ target (upper right) has the largest lobe, because it can still be perceived at fixation locations that are far removed from the location of this target. The ⊓ shaped target (lower right) and the ∧-shaped target are somewhat less distinct, resulting in intermediate lobe sizes. The small line segment (lower left) has a small visual lobe because it can not be distinguished from its local background when fixation is outside the small circle.

The conspicuity distance is closely related to the concept of conspicuity area, which is operationally defined as the peripheral area around the central fixation point from which specific target information can be extracted in a single glimpse (Engel, 1971, 1974, 1977). The conspicuity area is small if the target is embedded in a complex background (a surround with high feature variability) or if the target is surrounded by irregularly positioned nontargets of high similarity (a surround with high spatial variability). The conspicuity area is large if the target stands out clearly from a homogeneous background.

Procedure

Target conspicuity is measured as follows. The slide representing the scene and the target is projected continuously. First a moveable fixation dot is superimposed on the projected image by means of a laser pointer. This fixation dot is initially positioned at a large angular distance from the target location. Subjects are then instructed (i) to move the pointer slowly in the direction of the target while fixating the laser dot projected on the screen and (ii) to stop moving the pointer when the target first becomes noticeable in the peripheral visual field. The image is then replaced by the projection of a reference grid and the position of the fixation dot relative to this grid is recorded. Since the position of the target is known, the distance from the target at which its visibility is first reported can then be computed. This distance is adopted as the characteristic spatial extent of the conspicuity area of the target.

Each measurement is repeated at least 3 times. Subjects are usually able to make a setting within 15 seconds.

Results

The results will be discussed in Section 5, together with the results of the selected computational target distinctness measures.

4 COMPUTATIONAL TARGET DISTINCTNESS METRICS

4.1 Background

Various metrics have been developed to compute visual target distinctness from digital imagery. Visual distinctness metrics can be used to compare and rank target detectability, and to quantify background or scene complexity.

Some metrics are based directly on the sampled luminance values, others involve some (non)linear or (non)invertible transformation on the pixel values. They may be computed

locally	–	over the target area and its immediate surround,
semi-locally	–	on distinct locations in the scene), or
globally	–	over the entire scene.

Local metrics or signal-to-noise ratios quantify the distinctness of a target in its immediate surroundings. The general idea is that a target that is highly similar to its local background will be hard to see. *Semi-local metrics* are based on the calculation of likely fixation points for a human observer searching the scene for a target. The general idea is that a target will be hard to find (will be inconspicuous) if the inspection of the scene requires a large number of fixations. Fixation points are assumed to correspond to local extrema of

variance	-	of the graylevel distribution (Moravec, 1977),
busyness	-	(Burt, 1988a,b),
curvature	-	of the edge map (Lamdan et al., 1988; Yeshurun & Schwartz, 1989), or
symmetry	-	of the graylevel distribution (Reisfeld et al., 1990, 1995).
saliency	-	which may be any combination of the output of generalized difference operators operating on length, orientation, contrast, contour curvature, size, perimeter, and average graylevel (Milanese, 1993; Milanese et al., 1992, 1994).

Global conspicuity metrics or signal-to-clutter ratios take into account the overall structural composition of the scene. The general idea is that a target situated in a busy scene (a complex scene with a large amount of detail similar to the target, or a scene with a high structural variability) will be less conspicuous than the same target situated in a relatively empty scene (a scene with low variability).

Targets which are similar either to their local background or to many details in other parts of the scene are harder to detect than targets which are highly dissimilar to these structures. Also, the visual distinctness of a target decreases with increasing variability of the scene. Simple target distinctness metrics like the mean, median and standard deviation of the pixel values taken over the target support and part of its local background, do not capture these effects, and can therefore not predict human visual search and detection performance with highly resolved targets in complex scenes. It is found that detection performance depends mainly on the energy contrast between a target and its local background, whereas recognition depends mainly on the structural dissimilarity between a target and its surround (Braje et al., 1995; Caelli & Moraglia, 1986). The obscuring effect, resulting from the structural similarity between a target and its background, is called *clutter*. Since the concept is inherently elusive, attempts to quantify clutter have only been partly succesful (e.g. Cathcart et al., 1989; Shirvaikar & Trivedi, 1992; Schmieder & Weathersby, 1983; Tidhar et al., 1994; Waldman et al., 1988). For complex scenes, the spatial relationships (shape and relative location) of features in an image can have a greater effect on detection than the relative luminance of the features (e.g. Cathcart et al., 1989). Higher overall contrast may even reduce the amount of perceived clutter because confusing objects are more readily recognized for what they are — nontarget scene elements. The definition of clutter should account for this type of cognitive screening.

Current target acquisition models typically use first- or second order statistical metrics to describe the scene information content. First order metrics are only a function of pixel intensities and contain no information about relative pixel locations (spatial image structure). Second order metrics do contain some spatial information, but it is very difficult to determine how much and whether it is relevant to human spatial vision. Computational models of early human vision process an input image through various spatial and temporal bandpass filters and compute first order statistical properties of the filtered images to compute a target distinctness metric.

The rest of this section is the result of a literature study that was undertaken to investigate the different approaches to the computation of visual target distinctness that are currently available. The methods that have been reported to correlate best with human observer performance in visual search and detection tasks are selected for further evaluation in this pilot study (for the results of this evaluation see Section 5).

4.2 First Order Metrics

The shape of the first-order statistics or gray-level histogram of an image provides many clues as to the character of the image. A flat histogram corresponds to a low-contrast image. A bimodal histogram suggests that the image contains an object with a narrow amplitude range against a background with a different amplitude. Measures providing quantitative shape description of first-order histograms include the mean, standard deviation, skewness, kurtosis, energy, and entropy (Pratt, 1991). First order measures are inherently limited in their power to characterize perceptual differences in graylevel distributions (Shirvaikar & Trivedi, 1992). Examples of first order target distinctness measures are the following:

$$\Delta T = |\mu_t - \mu_b| \tag{1}$$

$$\Delta T_{rss} = \sqrt{(\mu_t - \mu_b)^2 + \sigma_t^2} \tag{2}$$

$$\Delta T_{rss4} = \sqrt{(\mu_t - \mu_b)^2 + 4\,\sigma_t^2} \tag{3}$$

$$\Delta T_{suma} = |\mu_t - \mu_b| + |\sigma_t - \sigma_b| \tag{4}$$

$$\Delta T_{sum} = |\mu_t - \mu_b| + \sigma_t \tag{5}$$

$$\text{Doyle} = \sqrt{(\mu_t - \mu_b)^2 + (\sigma_t - \sigma_b)^2} \tag{6}$$

$$\text{Doyle}_{\text{mod}} = \sqrt{(\mu_t - \mu_b)^2 + k\,(\sigma_t - \sigma_b)^2} \tag{7}$$

$$nrms = \frac{\sigma_{t+b}}{\mu_{t+b}} \tag{8}$$

where

t and b refer to respectively the target and background area, and
μ and σ represent respectively the mean and the maximum likelihood standard deviation of the graylevel distribution over the target or local background support.

The Doyle metric (6) and the modified Doyle metric (7) (with $k = 0.4$) have been reported to yield the highest correlation with psychophysically determined target distinctness measures (Copeland et al. , 1996). It has also been shown that search time progressively increases with the value of the $nrms$ metric (Kosnik, 1995). Hence, these metrics both appear suitable to rank order targets with respect to their visual distinctness.

4.3 Second Order Metrics

The second-order histogram or graylevel coocurrency (GLC) matrix represents the frequency of one graylevel occuring in a specified linear spatial relationship with another graylevel, within the area under investigation (Baraldi & Parmiggiani, 1995; Rosenfeld & Kak, 1982; Pratt, 1991). Hence, it embodies graylevel as well as pixel position information. The second-order histogram is defined as

$$P_\Delta(i,j) = \frac{1}{N}\sum_{k=1}^{N} f(x_k = i\ ,\ x_{k+\Delta} = j)\ \ , \tag{9}$$

where

$(x_k\,,\,x_{k+\Delta})$ represents a pair of pixels with respectively
(i,j) the graylevel values, $i,j \in [0, G-1]$, and separated by
$\Delta = (s,\theta)$ a polar displacement vector, with
s the distance between the pixels, and
θ the angle of the line through the pixles with respect to the image reference axis, and
$f(\cdot)$ $= 1$ if $x_k = i$ and $x_{k+\Delta} = j$
$= 0$ otherwise, and
G the number of graylevels in the image,
N the total number of pixels in the measurement window.

The GLC matrix is of dimension $G \times G$. If the pixel pairs within the image are highly correlated the entries in the GLC matrix will be clustered along its diagonal. Various measures have been proposed to specify the energy spread about the diagonal (Haralick et al. , 1976; Baraldi & Parmiggiani, 1995; Pratt, 1991; Trivedi et al. , 1984; for a recent overview see `http://www.cssip.elec.uq.edu.au/~guy/meastex/meastex.html`). Each of these measures captures some predominant visual characteristic of the image graylevel distribution. By computing these measures within a restricted window over the target and background area and defining a suitable distance measure it is possible to evaluate target distinctness. However, because of its large dimensionality it is difficult to use the GLC matrix in its raw form.

4.4 Normalised Clutter

Waldman et al. (1988) introduced a normalised clutter parameter that represents the amount of background texture that is similar to the target in size, shape and orientation. It is computed from the normalised joint amplitude histogram or coocurrence matrix that represents the transition probability between different graylevels at different locations in the image seperated by a vector of a given length and orientation. The amount of detail (texture) in the image determines the spread of this matrix about its main diagonal. For a scene consisting only of large details the entries of the matrix near the main diagonal will be large and those further from the main diagonal will be small. In the limit of a uniform scene there is only one non-zero entry in this matrix. The spread of the coocurrency matrix increases with the variability of the scene. The amount of clutter C is calculated as the mean of the relative texture size times the total distance weighted transition probability:

$$C = \frac{s}{T} B(\Delta) \tag{10}$$

where

s is the average texture element size,
T is the average target size in all directions,
Δ is the polar displacement.

The absolute value B is given by

$$B(\Delta) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i-j| P_{\Delta}(i,j) \ . \tag{11}$$

The normalised clutter value C_N is then defined as

$$C_N = \min\left(\frac{C}{B_E}, 1\right) \ , \tag{12}$$

where B_E is the expected value of B. This definition is such that clutter is an inherent property of the scene (target and background together), and independent of the sensor. It is symmetric with respect to target size and background texture size. It is always largest when the two are equal, but if either is twice the other, the clutter is reduced by the same amount. A uniform background always produces zero clutter by this measure, regardless of the target characteristics. A background densely packed with texture elements that are all the same shape, size and orientation as the target produces the maximum clutter of unity. The method works well for images with normally distributed synthetic backgrounds (Shirvaikar & Trivedi, 1992). For real images it only works if the texture element size is much smaller than the target size. No observer validation of this clutter metric is available. Since the assumption of normally distributed backgrounds is not realistic for complex natural images this method is not selected for further evaluation in the current pilot study.

4.5 Texture-based Clutter

The inertia of the GLC matrix is defined as (Pratt, 1991)

$$I(\Delta) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 P_\Delta(i,j) \ . \tag{13}$$

Shirvaikar and Trivedi (1992) proposed a global texture-based image clutter (TIC) measure that is based on this inertia measure. It is defined as

$$\text{TIC} = \frac{I(\Delta)}{\Delta} \tag{14}$$

where Δ is the characteristic target size. Like Waldman et al. 's (1988) normalised clutter measure (Eq. 12) the TIC measure also depends on target size. It provides zero weight to diagonal terms representing transitions between similar graylevels and progressively higher weights to off-diagonal entries representing transitions between pixels with dissimilar graylevels. In contrast to Waldman et al. 's normalised clutter measure the TIC measure utilizes all values in the GLC inertia matrix. However, it appears only slightly more succesful in capturing the perceptual meaningful information present in the GLC matrix (Shirvaikar and Trivedi, 1992). The authors suggest that additional measures are indeed required to include this information. Therefore, this measure is not further investigated in the present study.

4.6 Statistical Variance

Schmieder and Weathersby (1983) proposed to quantify the global amount of clutter in an image by

- dividing the image into blocks for which each linear dimension is twice the corresponding target dimension,
- calculating the gray level variance within each block, and
- taking the root mean square of the gray level variance over all blocks.

This metric, also called the statistical-variance metric (Rotman et al. , 1994a,b), is given by

$$SV = \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right)^{0.5} \tag{15}$$

where

N is the number of blocks into which the image is divided, and
σ_i^2 is the gray level variance within the i'th block, given by

$$\sigma_i^2 = \sum_{j=1}^{k} \frac{(X_{ij} - \mu)^2}{k} \tag{16}$$

where

X_{ij} is the gray level value of the j'th pixel in the i'th block,
μ is the average gray value in the i'th block, and
k is the total number of pixels in a block.

The signal-to-clutter ratio (SCR) of the image is then defined as the target gray level contrast divided by SV:

$$SCR = \frac{|\,\text{extremal target value - background mean}\,|}{SV} \quad . \tag{17}$$

This definition effectively introduces gray level normalization on a block by block basis. Hence, unlike the conventional gray level standard deviation, it avoids yielding a large clutter value for relatively uncluttered multimodal scenes. Signal and clutter have the same physical meaning, but both ignore the spatial structure of the graylevel distribution. A block of stripes and a "salt and pepper" block may have the same clutter value.

This clutter metric was checked on experimentally measured target detection probabilities. Good correlation between the detection probability and the value of the SCR was found for synthetic pictures characterising rural scenes (Schmieder and Weathersby, 1983). However, less satisfactory results were obtained when the metric was tested against target detection in urban synthetic scenes (Cathcart et al., 1989). For lower values of the SCR the targets were easier to find in urban scenes than in rural scenes. In urban scenes, the spatial relationships of features in the image had a greater effect on detection than the relative luminance of the features. This suggests that cognitive effects may have the largest effect on target conspicuity.

From visual search experiments it was found that the SV metric correlates rather weakly (a) with detection time (Birkmire et al., 1992) and (b) with human visual fixation behaviour (Rotman et al., 1994a). The SV metric is therefore not selected for further evaluation in the present pilot study.

4.7 Probability of Edge

Tidhar et al. (1994) introduced the Probability of Edge (POE) metric that uses the number of significant edge points as a measure of image clutter. The POE is determined as follows. First, the image is divided into blocks twice the apparent size of the target in each dimension. Second, a difference of offset Gaussians (DOOG) edge operator is applied to each block to simulate one of the channels in preattentive human vision. Third, the resulting edge image is thresholded and the number of points with a value above threshold T in the i'th block is computed as $\text{POE}_{i,T}$. The POE is then calculated as

$$\text{POE} = \frac{1}{N}\sum_{i=1}^{N} \text{POE}^2_{i,T} \tag{18}$$

and the corresponding SCR is defined as the Michelson contrast[3] of the target divided by the POE metric. The motivation for this approach is that preattentive vision is known to be drawn to edges (Caelli and Moraglia, 1986; Marr, 1982).

[3] The Michelson contrast C is defined as

$$C \;=\; \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}\ 100\%$$

with $L_{\max}$ and $L_{\min}$ respectively the maximum and minimum local luminance values.

The POE metric assumes that the *number* of edge points drives the search or fixation process, in contrast to the statistical variance measure SV (Eq. 15), which is sensitive to the *magnitude* of the edges, i.e. the intensity gradient at the edges. However, it relates to edge density only and does not represent the spatial structure of the graylevel distribution.

From visual search experiments it was found that the POE metric correlates rather weakly with detection time (Birkmire et al., 1992). In a comparative study Rotman et al. (1994a) found that in most cases the POE metric was inferior to other metrics in its capability to predict human visual fixation behaviour. Therefore, this metric is not studied any further in this report.

4.8 Circular Symmetry

This metric is based on the assumption that image areas of high symmetry and large gradients function as visual clutter. It has been described in great detail in Reisfeld et al. (1990, 1995). Here it suffices to say that every pixel P in the image is assigned a set of values $S_8(r, P)$ based on the local gradient and the symmetry in the r'th direction (where the radius index r ranges from 1 to 8 and the radial plane has been divided into 8 radial segments). The circular (overall) symmetry CS_8 at P is then given by

$$CS_8(P) \;=\; \prod_{r=1}^{8} (1 + S_8(i, P)) \quad . \tag{19}$$

Points of high values of CS_8 will have high symmetry. The procedure to computate the global value of the CS_8 metric is similar to the procedure that is used to compute the statistical variance metric SV (Eq. 15). First, the image is divided into N rectangular blocks. For each block i the sum $CS_{8,i}$ of the $CS_8(P)$ values for all pixels P in that block ($P \in i$) is computed:

$$CS_{8,i} \;=\; \sum_{P \in i} CS_8(P) \quad . \tag{20}$$

The global value of the CS_8 metric for the entire picture is then given by

$$CS_8 \;=\; \left(\frac{1}{N} \sum_{i=1}^{N} CS_{8,i}^2 \right)^{0.5} \quad . \tag{21}$$

Tidhar et al. (1994) defined a semi-local clutter metric as the number of likely fixation points that are within a given distance around the target, but not on the target. The larger this number is, the more distracted the observer will be and hence the less likely the observer is to find the target. The semi-local metric was evaluated on experimental search data obtained for targets in realistic aireal imagery. The fixation points were selected based on symmetry (Reisfeld et al., 1990, 1995). The mean detection time appeared to correlate with the semi-local metric. However, Rotman et al. (1994a) found that in most cases the CS_8 metric is inferior to other metrics in its capability to predict human visual fixation behaviour.

4.9 Peak Signal

The peak signal ΔT metric (Rotman et al., 1994) is based on the contrast between local extrema and their background. Again, the image is divided into N rectangular blocks. The user specifies a tolerance ΔT and a minimum cluster size. For each block, the algorithm starts with the pixel in the top left corner and moves to its right neighbour. If the difference in the pixel values is within the tolerance (selected empirically to be 0.7 times the average brightness intensity), the two pixel values are averaged to produce a 2 pixel cluster with a single value. If not, the neigbour pixel is assigned to a new cluster. The process is repeated until all pixels belong to a cluster. The peak signal (PS) ΔT metric for the i'th block is given by

$$\Delta T_{PS}(i) = \frac{T_{\text{max}} - T_{\text{min}}}{\dfrac{1 + |A(T_{\text{max}}) + A(T_{\text{min}})|}{A(T_{\text{max}}) - A(T_{\text{min}})}} \tag{22}$$

where

T_{max} is the largest cluster value in block i,
T_{min} is the smallest cluster value in block i, and
$A(\cdot)$ denotes the area corresponding to a cluster of pixels.

The numerator in Equation 22 measures the difference between the extremal cluster values in a block. Large values imply a high visual contrast and a large amount of clutter. This numerator is weighted by a denominator that varies between 1 and 2. For the limit when $A(T_{\text{max}}) = A(T_{\text{min}})$ the denominator is 1. However, when one of the clusters is very much larger than the other, its value is closer to 2. This weighting factor was introduced based on the subjective observation that clusters of drastically different areas appear to produce less subjective clutter than clusters with similar areas.

The global value of the ΔT_{PS} metric for the entire picture is then computed as

$$\Delta T_{PS} = \left(\frac{1}{N} \sum_{i=1}^{N} \Delta T_{PS}(i)^2 \right)^{0.5} . \tag{23}$$

In a comparative study Rotman et al. (1994a) found that in most cases the ΔT_{PS} metric was superior to other metrics in its capability to predict human visual fixation behaviour. However, the practical value of this measure is limited because it is computationally very expensive. This precludes for example the incorporation of this measure into wargames, where target distinctness needs to be updated every time step. Therefore, this method is not selected for further evaluation in the current pilot study.

4.10 Target Complexity

To determine the detectability of a target the pronouncedness of its edges should be taken into account. A target will be readily observable if it has well-defined edge points relative to its interior. In contrast, a scene with a wide distribution of edge strengths will be difficult to resolve.

Recently a clutter metric was introduced that is based on a Kolgomorov-Smirnov test on the cumulative distribution of (DOOG) edge points over the target and its immediate surround (a region about twice the size of the target in each dimension: Rotman et al., 1994a; Tidhar et al., 1994). Let the histogram of edge intensities be given by

$$\{N_i\}_{i=0}^{L-1} \tag{24}$$

where

0 and $G-1$ represent respectively the minimum and maximum of the histogram values, and

N_i is the number of points with value i.

The corresponding cumulative distribution is then given by

$$S_N(i) = \frac{1}{M^2} \sum_{j=0} iN_j \tag{25}$$

where

M^2 is the number of points in the histogram. with the following properties:

$S_N(i) = 0$ for $i < 0$,
$S_N(i) = 1$ for $i > G-1$,
$S_N(i) < S_N(i+1)$.

The target detectability is taken proportional to the mean absolute distance between the cumulative edge histograms of respectively the observed target section (S_N) and the null hypothesis (all values are equally likely or the uniform distribution) $P(i)$

$$TC = \frac{1}{G} \sum_{i=0}^{G-1} |S_N(i) - P(i)| \tag{26}$$

where

$P(i) = \sum_{j=0}^{i-1} \frac{1}{G} = \frac{i}{G}$.

The TC metric was evaluated on experimental search data obtained for human observers searching for targets in realistic aireal imagery (Rotman et al., 1994; Tidhar et al., 1994). In some situations it shows a reasonable overall correlation with the mean detection time. However, the metric fails completely in situations where the target is the most salient item in a particular sector even though the rest of the image is crowded. Because of its limited applicability this metric is not studied any further in the current pilot study.

4.11 Complex Contrast

Lillesæter (1993) suggested that the perceived contrast of a structured target on a complex background can be represented by a measure that effectively combines (a) the mean luminance contrast of the target and background area with (b) the average of the luminance variations along the target/background borderline. The complex contrast K thus defined is given by K:

$$K = a\left|\overline{\ln G_t} - \overline{\ln G_b}\right| + \frac{b}{Z}\oint_Z \left|\ln\left(\frac{G_t}{G_b}\right)\right| \tag{27}$$

where

G represents the pixel grayvalue (sampled luminance) distribution,
t , b refer to respectively the target and background area,
Z is the outer target contour, and
a , b are weight factors that sum to unity, both tentatively set to 0.5.

The first term on the right-hand side corresponds to the mean area contrast. The integration in the second term is carried out along the entire length Z of the outer target contour. There is no natural limitation of the portion of the background that should be taken into account. The dilation of the target area with a square structuring element with the size of the target itself is (rather arbitarily) adopted as the local target background area. It is assumed that a possible extension of this area does not influence the resulting contrast unless the luminance of the bacjkground changes significantly. This complex contrast measure has been implemented in the US Night Vision Laboratory Static Performance Model for Thermal Viewing Systems (Skjervold, 1995). No psychophysical validation studies have been published so far. Because of its simplicity and relevance to human vision this complex contrast measure is included in the present comparative pilot evaluation study.

4.12 Oriented Filter Models

This subsection serves as an introduction to the vision models that will be discussed in the rest of this section. It presents the rationale of the approach and some principles that are common to these models.

Background

Over the years, a large amount of research has been directed toward the capability of the human visual system to detect arbitrary targets in complex scenes (Carlson & Cohen, 1980; Marr, 1982; Caelli & Moraglia, 1986). The properties of the excitatory and inhibitory receptive fields that are arranged in concentric circles in the retina, and of the receptive fields that are sensitive to particular orientations in the striate cortex have all been extensively studied. This knowledge of the human visual system is widely used to construct early vision models (Gerhart et al., 1995; Watson, 1983; Witus et al., 1995a,b), image distortion metrics (Ahumada & Beard, 1996; Fdez-Vidal et al., 1996a,b; Heckmann et al., 1995; Lakshmanan et al., 1995; Martinez-Baena, 1996; Meitzler et al., 1995; Witus et al., 1995b), and models of human visual target acquisition (Gerhart, 1995; Doll et al., 1993; Overington, 1982; Witus et al., 1995c).

Image features and oriented energy

Many different computational approaches to detection of features in an image have been developed, including first-, second- and third-derivative techniques, surface fitting and mathematical morphology. Psychophysical research indicates that features are perceived at those points in an image where the Fourier components of the image function are maximally in phase (Burr & Morrone, 1990; Morrone et al., 1986; Morrone & Burr, 1988).

Phase congruency is a rather awkward quantity to calculate. However, Venkatesh and Owens (1989) have shown that the phase congruency function is directly proportional to the local energy function. Points of maximum phase congruency therefore coincide with peaks in the local energy function. Ronse (1995) provided a mathematical framework for this approach to feature detection and extended the method to two-dimensional non-periodic images.

Based on these results a local energy visual feature detector was proposed (Morrone & Owens, 1987; Morrone & Burr, 1988; Owens et al., 1989). In this scheme the local energy of an image is computed as the norm of a vector whose components are the image intensity and its Hilbert transform.

Let $F(x)$ represent the image intensity function of a one-dimensional windowed luminance profile, and let $H(x)$ represent the Hilbert transform of $F(x)$. In terms of a Fourier expansion $F(x)$ and $H(x)$ can be written as

$$\begin{aligned} F(x) &= \sum a_n \sin(n\omega x + \phi_n) \quad , \quad \text{and} \\ H(x) &= \sum a_n \cos(n\omega x + \phi_n) \quad , \end{aligned} \tag{28}$$

where $a_n > 0$, ϕ_n is the phase shift of the n^{th} terms and the summation is taken over the non-negative integers.

Let **E** be the analytic signal (Bracewell, 1965) given by

$$\begin{aligned} \mathbf{E}(\mathbf{x}) &= F(x) - i\,H(x) \\ &= \sum a_n\,(\,sin\,(n\omega x + \phi_n) - i\,cos\,(n\omega x + \phi_n)) \\ &= -i \sum (\,a_n\,(cos\,(n\omega x + \phi_n) + i\,sin\,(n\omega x + \phi_n))) \\ &= e^{\,i\,(-\pi/2+2k\pi)}\,(H(x) + iF(x)) \quad , \end{aligned} \tag{29}$$

where i represents the square root of -1. The local energy function $E(x)$ is then defined as the norm of the analytic funtion **E**:

$$E(x) = |\mathbf{E}(\mathbf{x})| = \left(F(x)^2 + H(x)^2\right)^{0.5} \quad . \tag{30}$$

E is a vector sum of an infinite number of components. If these components are distributed with a small standard deviation around the mean phase value, the norm of **E** will attain a local maximum. Points in space where this condition occurs are termed points of maximum phase congruency. Extensive experimentation has established that image features coincide with the local maxima of **E**, and therefore with the points of maximum phase congruency (Kovesi, 1995; Morrone & Burr, 1988; Morrone & Owens, 1987; Owens et al., 1989; Venkatesh & Owens, 1990).

Rather than compute the local energy via the Hilbert transform of the original luminance function one can calculate a measure of local energy by convolving the image with a pair of filters in quadrature. One of these filters is designed to remove the DC component. The other one is in quadrature with the first. Application of these filters to the image results in two band-pass filtered versions of the image, one being 90° phase shift of the other. The local energy is then computed as the square root of the sum of squares of these two convolutions. This alternative energy function $\hat{E}$ is thus defined as

$$\hat{E}(x) = \Big((S * F(x))^2 + (A * F(x))^2 \Big)^{0.5} \tag{31}$$

where

$F(x)$ represents the original intensity distribution,
$S(x)$ is a symmetric mask,
$A(x)$ is an anti-symmetric mask, and
$*$ denotes the convolution operation.

The masks are usually chosen such that they have zero mean, identical L^2 norm, and that they are orthogonal. The maxima of $\hat{E}$ then coincide with the images features. Typical three-point even and odd masks are respectively

$$\begin{aligned} S(x) &= \begin{array}{|c|c|c|}\hline -1 & 2 & -1 \\ \hline \end{array} \\ A(x) &= \begin{array}{|c|c|c|}\hline -1.73 & 0 & 1.73 \\ \hline \end{array} \end{aligned} \tag{32}$$

The local maxima of the generalized function $\hat{E}$ are then selected as the feature points in the image.

Since it is not possible to construct rotationally symmetric odd-symmetric filters the feature detection technique is usually extended to two dimensions by convolving the image with uni-dimensional masks in two orthogonal directions. The image features are then detected as the union of the local maxima from each uni-dimensional operation:

$$\chi(x, y) = \max\left(H_{\max} E_h(x, y), V_{\max} E_v(x, y)\right) \tag{33}$$

where

$\chi(x, y)$ = 1 if there is a feature (local energy maximum) at (x, y),
= 0 if there is no feature at (x, y),
max is the pairwise maximum operator (the binary union),
$E_h(x, y)$ is the horizontal energy function,
$E_v(x, y)$ is the vertical energy function,
$H_{\max}$ is the horizontal local maximum operator,
$V_{\max}$ is the vertical local maximum operator.

More elegant and theoretically better established approaches analyse two-dimensional images by employing oriented Gabor or wavelet filters (Kovesi, 1995; Robbins & Owens, 1994; Ronse, 1995). The calculation of energy from spatial filters in quadrature pairs is now central to many models of human visual perception (e.g. Adelson & Bergen, 1985; Heeger, 1987, 1988; Watson & Ahumada, 1985).

4.13 The Georgia Tech Vision Model

The Georgia Tech Vision Model (GTV) is a simulation of human visual search and detection (Doll et al., 1993, 1995). It incorporates findings from recent psychophysical and neurophysiological research on low-level visual processes, visual search, selective attention, and signal detection theory. It calculates a distinctness value for each location in the visual field. The probability that focal attention will be directed to a given object or location in the visual field is taken directly proportional to the local distinctness value. In GTV, distinctness depends on luminance contrast, chromatic contrast, temporal modulation and texture differences. A "pattern perception" module simulates the pattern detection capability of the visual system. It is based on recently reported two-stage models of visual texture segregation (Graham, 1991; Graham et al., 1992; Wilson et al., 1983).

The *first stage* of the pattern perception module filters the image by 24 spatial frequency and orientation selective filters (6 spatial frequency bands in 4 orientations, implemented as directional derivatives of DOG's; Wilson et al., 1983). The filtered images are then rectified (by taking the absolute value) to obtain measures of contrast energy in each channel. High-frequency artifacts introduced by the rectification step are removed by low-pass filtering. The first stage is concluded by adjustment of the filter gains to simulate effects of

- adaptation to the mean luminance of the image (Kelly, 1972),
- orientation dependency of pattern sensitivity (Campbell et al., 1966; Mitchell & Wilkinson, 1974), and
- interactions among spatial frequency channels (Greenlee & Magnussen, 1988).

The *second stage* of the pattern perception module simulates the capability of the visual system to segregate regions with different visual textures (Wilson & Richards, 1992). Each output channel (image) resulting from the first stage is band-pass filtered with a filter whose upper cut-off is approximately one octave lower than that of the first stage channel. The resulting images are rectified and low-pass filtered to remove artifacts introduced by rectification.

A weighted vector summation (Quick-pooling: Quick, 1974) is used to pool the resulting 24 channels. Depending on the nature of the visual search task an observer may attend selectively, or more heavily, to certain channels (e.g. observer expectations). The weights used in the vector summation simulate this top-down processing. The total of the graylevel values of the resulting (pooled) image is taken as a global clutter measure.

The GTV clutter metric is highly correlated with the SV metric (Eq. 15: Doll et al., 1993, 1995).

When this study was performed the GTV model was being evaluated by the U.S. Army Material Systems Analysis Activity (AMSAA; at Aberdeen Proving Ground, Maryland, USA) and unfortunately not available for evaluation by third parties.

4.14 The TARDEC Vision Model

The TARDEC Vision Model (TVM) emulates human visual search and detection (for a detailed description of the model see: Witus et al., 1995a). It transforms a digital image through various spatial and temporal filters, representing the stream of processing in the human visual system, beginning with the optics of the ocular media, and continuing with the retina, the lateral geniculate nucleus of the midbrain, and the visual processing areas of the occipital cortex. The transformed image representation is used to calculate a distinctness metric for a given target in the digitised input scene. The metric is used to predict human visual acquisition performance.

TVM captures the target distinctness metric in two stages. First, it computes a signal-to-noise ratio (SNR) for each individual channel indexed by temporal, luminance/color opponent, spatial scale, and orientation dimensions. The SNR is defined as the root-mean-square (RMS) modulation amplitude due to the target, normalised to the RMS of internal noise and background clutter. The clutter measure is the standard deviation of the RMS amplitude due to the local background signal. The internal noise term represents the differential sensitivity on each channel under different luminance adaptation conditions. The RMS of the SNR's for all visual channels weighted by the relative density of receptive fields in each channel is taken as the target distinctness metric.

TVM has been calibrated relative to an extensive set of laboratory human observer measurements (Gerhart et al., 1995). It has succesfully been applied (i) to predict subject responses to a variety of realistic stimuli under simulated driving conditions (Witus et al., 1995b), (ii) to evaluate observer performance with infra-red systems (Lakshmanan et al., 1995), (iii) to optimize the spatial configuration of the center high-mounted rear stoplight in a car brakelight configuration (Heckmann et al., 1995; Meitzler et al., 1995).

When this study was performed TVM was also being evaluated by the U.S. Army Material Systems Analysis Activity (AMSAA; at Aberdeen Proving Ground, Maryland, USA).

4.15 The Cortex Transform

Ahumada (Ahumada et al., 1995; Ahumada & Beard, 1996) developed a target distinctness metric that is based on the Cortex transform of Watson (1983, 1987). The model compares two input images: one depicting a scene with the target for which the distinctness needs to be evaluated and a second one, depicting exactly the same scene without the target. When the latter image is not available it can artificially be produced by replacing the target with approriate background imagery.

The model calculations involve the following steps (for an extensive description of the approach see: Ahumada & Beard, 1996). First, the images are converted to luminance contrast by subtracting and then dividing by the background image mean luminance. Second, a contrast sensitivity filter is then applied to both images. Third, the Cortex transform is then applied to both images. Fourth, in the Cortex transform domain, the differences between the transforms of the object and background images are then divided by the absolute value of the corresponding coefficient from the background image to the 0.7 power if that absolute value is greater than one (threshold). Fifth, the absolute value of these scaled Cortex transform coefficient differences are raised to a power, summed, and then taken to the inverse power. For the case that the exponent is infinity, the maximum absolute difference is computed.

The contrast sensitivity filters are calibrated separately for the summation exponents 2, 4, and infinity. They were designed to fit the prediction of Barten's contrast sensitivity formula (1993) for 1.33 deg square grating patches at five spatial frequencies centered in each of the five bandpass channels of the multiple channel model.

The model has succesfully been applied to predict the relative detectability of military vehicles in natural backgrounds (Ahumada et al., 1995), and the detection of aircraft in noisy images (Ahumada & Beard, 1996).

4.16 Perceptual Distortion

The Computational Science Group at the University of Granada (Spain) recently investigated the capability of several root mean square error metrics to predict visual target distinctness (Fdez-Vidal et al., 1996a,b; Martinez-Baena et al., 1996).

Martinez-Baena et al. (1996) introduced a method to calculate a perceptual distortion measure between two complex images. In this approach, one image serves as the reference signal while the other image is considered to be a distorted version of this reference signal.

This method works as follows. First, the Fourier transform of the reference image is computed. Second, the resulting reference spectrum is partitioned into a set of orientation and spatial-frequency selective filters. The receptive fields of these filters correspond to Gabor functions. Third, the sensors that respond most strongly to significant orientation and spatial-frequency components of the reference image are identified and their corresponding receptive field parameters are obtained. Fifth, the reference image and the input image are convolved with the resulting set of selected filters. Finally, the distortion between the pair of complex images is computed as a function of the differences of the signal activity of corresponding filters.

Two different distortion metrics are proposed. The first metric d_1 is based on a weighted sum of the squared absolute differences between corresponding filters outputs, weighted by the respective average graylevel of the image. The second metric d_2 computes a weighted sum of the squared absolute differences between the corresponding normalised power spectrum coefficients. Both metrics correlate well with the subjective ranking of human observers judging (i) the perceptual image quality of reconstructions of JPEG compressed images, and (ii) the visibility of targets in images with different amounts of noise (Martinez-Baena et al., 1996)

The method can be applied to evaluate target distinctness by adopting respectively (a) the scene with the target as the distorted signal, and (b) exactly the same scene without the target as the reference signal.

In a follow-up study Fdez-Vidal et al. (1996a) compared the capability of three different root mean square error measures of the difference between the target scenes (Fig. 7) and the corresponding empty scenes (Fig. 8) to quantify the visual distinctness of the targets in Figures 1–6 as perceived by human observers. The error measures were not simply computed globally over the entire image support, but semi-locally at locations that are likely to attract human fixation (i.e. that are likely to be inspected because they are seen as characteristic features). The local regions chosen to compute the error metric corresponded to neigbourhoods of

- Laplacian zero crossings,
- points of maximal phase congruency, and
- points of maximal energy of the most activated sensors,

for the reference image. The last two metrics (maximal phase congruency and maximal energy) yield a target distinctness rank order that correlates with human observer performance, whereas the first measure (based on Laplacian zero crossings) fails to produces a correct rank order (Fdez-Vidal et al. , 1996a).

Recently Fdez-Vidal et al. (1996b) showed that a perceptual root mean square distortion metric, that effectively *combines* differences in graylevel values, phase congruency values, and phase of local energy values, computed over small local neigbourhoods of points of maximum phase congruency, yields a target distinctness rank order identical to the one found for human observers. The restriction of all computations to locations near points of maximum phase congruency is an attempt to simulate human selective attention, since previous studies have shown that these points correspond to features in the image that are most likely to attract fixation (points that are likely to be inspected). This approach works as follows.

The original target scene is represented by an ensemble $\mathcal{T}$ of three seperable image representations

$$(T_j(x,y))_{(x,y)\in R} \quad ; \quad j = 1,2,3 \tag{34}$$

where

T_1 represents the graylevel values,
T_2 represents the corresponding phase congruency values,
T_3 represents the local energy, and
x, y represent the pixel coordinates.

Similarly, the corresponding empty scene is represented by three matrices $(E_j(x,y))$; $j = 1,2,3$ corresponding to the graylevel, phase-congruency, and phase of local energy values of pixels in the image domain R respectively.

The set of evaluation (fixation) points $W_{pc}(\mathcal{T})$, defined as the points of maximum phase congruency for the graylevel representation of the target scene $\mathcal{T}$, is given by

$$W_{pc}(\mathcal{T}) = \{ \, (x,y) \in R \mid (x,y) \text{ is a point of maximum phase congruency for } \mathcal{T} \} \ . \tag{35}$$

The perceptual distortion metric $d(\mathcal{T},\mathcal{E})$ is then defined as

$$d(\mathcal{T},\mathcal{E}) = \iint\limits_R \operatorname{dist}(\, \mathcal{T}(x,y) \, , \, \mathcal{E}(x,y) \,) \, \psi_{\mathcal{T}}(x,y) \, dx \, dy \tag{36}$$

where dist $(\mathcal{T}(x,y), \mathcal{E}(x,y))$ defines a normalised distance measure for the integral representations

$$\mathcal{T}(x,y) \quad = \quad (T_1(x,y), T_2(x,y), T_3(x,y)) \quad , \quad \text{and}$$

$$\mathcal{E}(x,y) \;=\; (E_1(x,y), E_2(x,y), E_3(x,y)) \;\;,$$

given by

$$\mathrm{dist}\,(\,\mathcal{T}(x,y)\,,\,\mathcal{E}(x,y)\,) = \sum_{i=1}^{3} \left[\frac{(\,T_i(x,y) - E_i(x,y)\,)^2}{\max_{(x,y)\in W_{pc}(\mathcal{T})} \left\{ (\,T_i(x,y) - E_i(x,y)\,)^2 \right\}} \right]^{0.5} \tag{37}$$

and the spatial sensitivity function $\psi_{\mathcal{T}}$ is given by

$$\psi_{\mathcal{T}} = \begin{cases} \dfrac{1}{\mathrm{Card}\,[W_{pc}(\mathcal{T})]} & \text{if } (x,y) \in W_{pc}(\mathcal{T}) \;\;, \\ 0 & \text{otherwise} \;\;. \end{cases} \tag{38}$$

4.17 Histogram Intersection

Background

In visual search humans typically perform a sequential foveation of highly selective regions of the scene. During fixation the foveated regions are analysed in detail. The foveated regions correspond (a) to locations that are either sufficiently distinct from their immediate surroundings to attract attention, or (b) to locations that are a priori likely to contain the items that are searched.

A salient target can attract attention even when it is not fixated. In this case the target is depicted in the peripheral visual field where the resolution is relatively low. To be effective as an attention attractor the feature difference giving rise to target distinctness therefore has to be resolution independent.

Contrast and color differences do not depend critically on resolution and viewing angle. Shape and texture differences are highly resolution and viewpoint dependent. Histograms are invariant to translation and rotation about the viewing axis, and change only slowly under change of angle of view, change in scale and occlusion.

The computation of target saliency involves the quantification of perceptual differences of the most general type between local regions of complex images. It appears that the local spatial information content in an image is to a large extent characterized by the first-order statistics of the local gray value distribution (i.e. the rank ordered gray value distribution, or, equivalently, the local gray value histogram: Lowitz, 1983, 1984). Differences between image regions can therefore often be detected by comparing their gray value histograms (e.g. Zamperoni, 1995). The analysis can be further refined by comparing e.g. the individual histograms of oriented filters of different pass-bands or colour histograms.

The image histogram

The image gray value histogram is a table that lists for each possible pixel (graylevel- or RGB-) value the number of pixels in the image that actually have that value. The domain of the histogram is the set of all possible pixel values. The range of the histogram is the set of positive integers ranging from 0 to the total number of pixels in the image. The sum of all entries in the histogram equals the total number of pixels in the image. Let $h(v)$ denote the histogram entry for value v, and let the image represent a luminance function quantised to 8 bits on a rectangular array of width N_w and height N_h. Then

$$\sum_v h(v) = N_w \times N_h$$

with

$$v \in [0, 255] \ .$$

The histogram may be computed over the entire image or over an arbitrarily shaped region of interest. If n_a is the number of pixels in the area over which the histogram is computed then

$$\sum_v h(v) = n_a \ .$$

The normalized histogram $H(v)$, given by

$$H(v) = \frac{h(v)}{n_a}$$

lists for each possible pixel value the fraction (percentage) of pixels in the image that actually have that value. $H(v)$ is analogous to the probability density function of statistics and may be considered as the probability of a pixel having value v. In this case

$$\sum_v H(v) = 1 \ .$$

Histogram matching

Image regions that appear visually similar should have similar normalized histograms. Note that the reverse statement need not be true (regions with similar histograms may look very different indeed). Corresponding bins of similar normalized histograms have approximately the same degree of occupancy. This suggests that histograms can be compared (matched) by computing their common part or intersection. The intersection of a target and background histogram tells what fraction of the target pixels is also found in the background. Histogram matching has succesfully been used to compute saliency maps to guide the focus of attention in automatic target detection algorithms and for model matching (Ballard & Brown, 1992; Ennesser & Medioni, 1995; Mel, 1996; Swain & Ballard, 1991).

Let H_T and H_B respectively represent the normalized histograms of a target and its local background, each containing n bins. The intersection of H_T and H_B is defined as the cumulative sum of the pairwise minimum of corresponding histogram entries

$$H_T \cap H_B = \sum_{v=1}^{n} \min \{H_T(v), H_B(v)\} \ .$$

The value of the intersection is between 0 (no histogram overlap) and 1 (complete overlap).

4.18 CAMAELEON

CAMAELEON is a computer model developed for the assessment of camouflage using digital image processing techniques (Hecker, 1992). The model requires digitised (color or graylevel) images for input. The user is required to manually delineate the target and background areas in the input image, and to specify the target area (m^2). CAMAELEON convolves the input image with a bank of bandpass (polar quadrature mirror) filters. From the filter outputs it computes local energy (contrast), local spatial frequency and local orientation. Local energy is computed as the sum over all bands of the energies of the individual filters. Local spatial frequency is computed as the vector sum of complex frequency over all bands. Local orientation is computed as the vector sum of directions over all bands. A normalized histogram can be obtained by dividing each entry of the histogram by the total number of pixels. The resulting entries represent the fraction of the total number of pixels that are in a certain state. CAMAELEON computes normalized histograms of

- the local energy (HE),
- the local orientation (HO), and
- the local frequency (HF)

for

the target area	:	(HE_T,HO_T, HF_T), and
the background area	:	(HE_B,HO_B, HF_B) .

CAMAELEON defines the degree of camouflage (target distinctness or signal-to-noise ratio) C as

$$C \ = \ (HE_T \cap HE_B) \cdot (HO_T \cap HO_B) \cdot (HF_T \cap HF_B) \tag{39}$$

with $C \in [0, 1]$. $C = 1$ corresponds to a target which has energy, orientation and frequency histograms that are identical to the corresponding local background histograms. As a result, the target may be hard to distinguish from its background (it is well camouflaged). $C < 1$ corresponds to a target which differs from its local background, either in energy, or in orientation, or in frequency, or in a combination thereof. The target has therefore some visual contrast with its local background and is probably visually detectable.

CAMAELEON predicts the detection range (km) for the target from an emperically derived function that relates target distinctness to probability of detection.

4.19 Edge distance metric

Rationale

The motivation for the edge distance metric, which will be introduced in this section, is that preattentive vision is known to be drawn to edges (Caelli and Moraglia, 1986; Marr, 1982). As a result, detection performance depends mainly on the energy contrast between a target and its local background. However, recognition depends mainly on the structural dissimilarity between a target and its surround (Braje et al., 1995; Caelli & Moraglia, 1986). The detection of targets of low visibility probably depends on the perception of

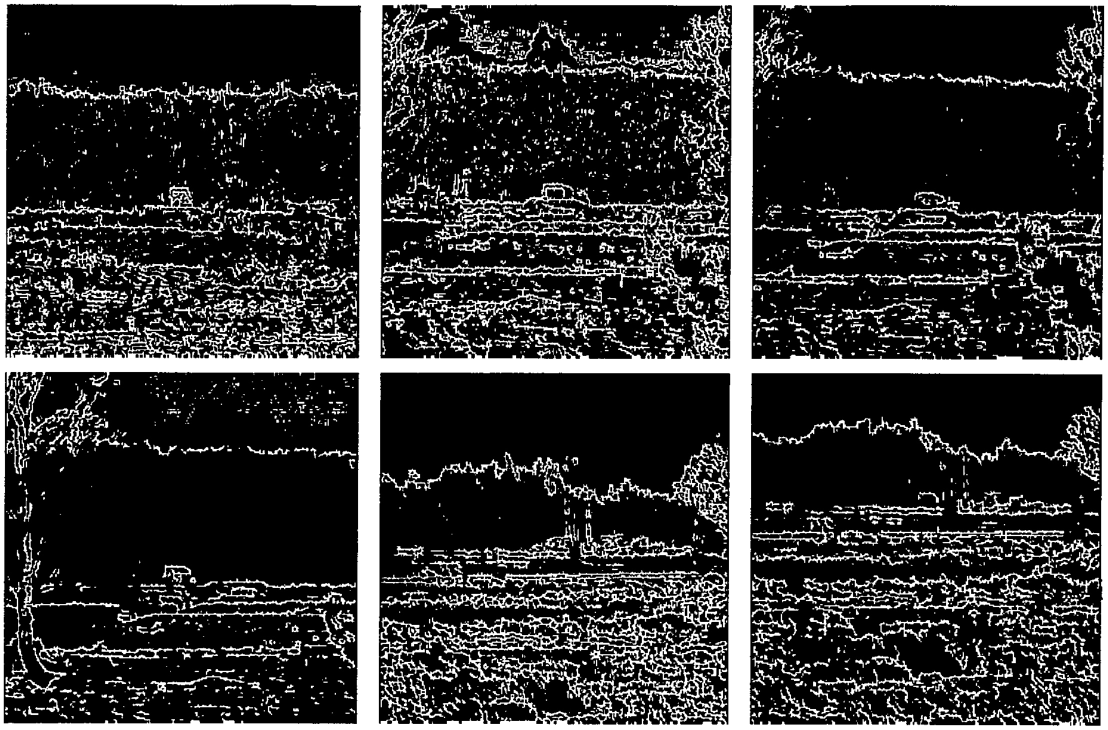


Fig. 12 Edge representation of target sections from Figs. 1-6 (from upper left to lower right), as shown in Fig. 7.

local structural gradients. The edge distance metric introduced in this section is an attempt to capture both the edge density and the structural composition of the scene in a single local measure.

Similar to the POE metric, the edge distance metric assumes that the search or fixation process is fully driven by the *density* of edge points in the scene. Therefore, information about the intensity gradient at the edges is discarded (unlike e.g. the statistical variance measure SV, given by Eq. 15, which is sensitive to the *magnitude* of the edges). In contrast to the POE metric, which relates to edge density only, the edge distance metric also represents the spatial structure of the graylevel distribution.

Before introducing the edge distance metric, the next two subsections first define the edge transform and the distance transform that are used to compute the new metric.

Edge Transform

Local discontinuities in image luminance are called *luminance edges*. Edges can be detected by applying a local difference or gradient operator to the image luminance function and thresholding the output (Rosenfeld & Kak, 1982). The result of this transformation is a cartoon-like image, representing the locations in the original image where the luminance gradient is strong (e.g. Fig. 12).

A wide range of edge detectors has been defined. Each of these detectors has its own characteristic performance (e.g. Pratt, 1991). Since the exact nature of the edge detector is irrelevant in this study, a simple second order or Laplacian difference operator is applied (Pratt, 1991, p. 519; Rosenfeld & Kak, 1982). The edge Laplacian at each location is obtained by convolving the image with the following gain-normalized, four-neighbour kernel

$$G = \frac{1}{8}\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} .$$

The output of edge detectors is usually corrupted by noise effects. As a result, many "false" edge points are detected which do not lie on border regions, and at many border points no edges are detected. There are many techniques to improve the output of edge detectors by deleting noise responses (thresholding techniques), filling edge gaps (edge linking) and enhancing the articulation or pronouncedness (edge thinning, see: Pratt, 1991; Rosenfeld & Kak, 1982).

In this study the noise in the Laplacian filtered images was reduced by thresholding them at a pixel value of 6/8. The result of this operation is a binary image in which all pixels with an edge strength greater than or equal to 6/8 are set to 1, and all pixels with an edge strength smaller than 6/8 are set to 0. The threshold value 6/8 was determined by eye and yields the best representation of the target contour with a minimal amount of noise. The result of this operation on the target sections from Figures 1–6 is shown in Figure 12.

Each of the abovementioned noise cleaning methods can be extremely useful as an image preprocessing operation in an implementation of the target conspicuity computation method proposed here. However, since this study only investigates the feasibility of the proposed method, these methods will not be investigated here. Next to the computational simplicity, an additional advantage of the use of a simple second order operator for edge detection in the present context is the fact that it provides a kind of worst case test. Hence, if the outcome of the proposed computational scheme correlates with results from human visual search and detection experiments, the scheme can probably be further improved by incorporating and tuning the abovementioned improvements.

Distance Transform

A distance transformation converts a binary digital image, consisting of feature (foreground) and non-feature (background) pixels, into a greylevel image in which each pixel has a value that corresponds to the distance to the nearest feature pixel. Distances computed by sequential algorithms that approximate global distances in the image by propagating local distances (i.e. distances between neigbouring pixels) over the image plane are also known as "chamfer" distances (Paglieroni, 1992).

Because of the discrete nature of digital images and the influence of noise on the location of edge points, it is an unnecessary waste of effort to compute exact Euclidian distances (Daniellson, 1980) from the inexact boundary pixels. In most digital image processing applications it is therefore preferable to use integers to represent distances. The two different local distances in a 3 × 3 pixel neighbourhood are the distance between the

$$\begin{array}{ccc} 7 & 5 & 7 \\ 5 & 0 & 5 \\ 7 & 5 & 7 \end{array}$$

(a)

$$\begin{array}{ccc} & 0 & 5 \\ 7 & 5 & 7 \end{array}$$

(b)

$$\begin{array}{ccc} 7 & 5 & 7 \\ 5 & 0 & \end{array}$$

(c)

Fig. 13 Masks for the parallel (a) and the sequential (backward and forward - b and c) chamfer-5/7 distance transform.

horizontal/vertical neighbours and between diagonal neighbours. The chamfer-5/7 distance (Borgefors, 1986) uses respectively the values 5 and 7 to represent these distances. This distance has a maximum difference of 4.21% from the true Euclidian distance (Borgefors, 1986; Verwer, 1991).

The chamfer-5/7 distance transform can be calculated sequentially by a two-pass algorithm as follows (see also Fig. 14). First, the distance image is initialized by attributing each feature pixel a distance zero and by setting the value of each non-feature or background pixel to infinity. A given distance transform is characterized by a mask whose entries are the local distances that are propagated over the image (Fig. 13a). In the sequential algorithm, the mask is separated into two masks (Figs. 13b,c). These (smaller) masks are passed over the image once each: first "forwards", from left to right and from top to bottom (Fig. 13b), and then "backwards", from right to left and from bottom to top (Fig. 13c). The origin of the mask is placed over each pixel in the original image. The local distance in each mask entry is added to the value of the image pixel directly "under" it. The new value of the central pixel is the minimum of these sums. The effect of this transform is illustrated in Figure 15.

Edge-distance metric

The *edge-distance target distinctness* measure D is defined as the ratio of the maximal distance value that is observed in the local background of the target and the maximal distance value that occurs within the target support:

$$D = \frac{d^{max}_{\text{background}}}{d^{max}_{\text{target}}} \ , \tag{40}$$

where d represents is the local distance value. D is large when there is not much structure in the image region surrounding the target (i.e. when there are large distances between the edges in the background), or when the edge density over the target support is high (i.e. when the mean distance between the edges over the target support is very low). D decreases with an increasing amount of detail surrounding the target and with a decreasing pronouncedness (level of detail) of the target itself. The rationale for this definition is that a target stands out (its visual distinctness is high) when it is situated in a homogeneous and relatively structureless area, whereas it is hard to find when it is located in a crowded environment.

```
Procedure FORWARD
for $i \leftarrow 2\,(1)$ rows
    for $j \leftarrow 2\,(1)$ columns $- 1$
        $v(i,j) \leftarrow \min\{v_{i-1,j-1}+7\,,\ v_{i-1,j}+5\,,$
                $v_{i-1,j+1}+7\,,\ v_{i,j-1}+5\,,\ v_{i,j}\}$
    end for
end for
```

```
Procedure BACKWARD
for $i \leftarrow$ rows $- 1\,(1)\,1$
    for $j \leftarrow$ columns $- 1\,(1)\,2$
        $v(i,j) \leftarrow \min\{v_{i+1,j+1}+7\,,\ v_{i+1,j}+5\,,$
                $v_{i+1,j-1}+7\,,\ v_{i,j+1}+5\,,\ v_{i,j}\}$
    end for
end for
```

Fig. 14 Algorithm for the sequential chamfer-5/7 distance transform.

```
 *  *  *  *  *  *  *      *  *  *  *  *  *  *     21 19 17 15 17 19 21
 *  *  *  *  *  *  *      *  *  *  *  *  *  *     19 14 12 10 12 14 19
 *  *  *  *  *  *  *      *  *  *  *  *  *  *     17 12  7  5  7 12 17
 *  *  *  0  *  *  *      *  *  *  0  5 10 15     15 10  5  0  5 10 15
 *  *  *  *  *  *  *      *  *  7  5  7 12 17     17 12  7  5  7 12 17
 *  *  *  *  *  *  *      * 14 12 10 12 14 19     19 14 12 10 12 14 19
 *  *  *  *  *  *  *     21 19 17 15 17 19 21     21 19 17 15 17 19 21

        (a)                      (b)                      (c)
```

Fig. 15 The computation of the chamfer-5/7 distance transform. (a) The original image with one feature point in the middle. (b) and (c): Results of respectively the forward and backward pass of the sequential chamfer-5/7 distance transform.

Most computational target distinctness metrics require either

- a precise delineation of the local background (e.g. the Camaeleon model: Hecker, 1992), or
- the definition of a zone within which background elements are able to interfere with target details (e.g. Tidhar et al., 1994), or
- images of the same scene both with and without the target (e.g. the TVM model: Witus et al., 1995a; and some image difference metrics based on early vision models: Ahumada & Beard, 1996; Fdez-Vidal et al., 1996a,b; Martinez-Baena et al., 1996).

The edge-distance target distinctness metric has neither of these requirements. Although Equation 40 is calculated over the target area and a restricted local background area, remote details still contribute to the target distinctness. The use of a predefined (fixed) local background area may be replaced by a procedure in which the distance image is first segmented (e.g. by using its watersheds: Serra, 1982).

The capability of the edge-distance metric to rank order targets with respect to their visual distinctness is evaluated in Section 5.

5 EVALUATING THE TARGET DISTINCTNESS METRICS

Table III lists both the outcome of the psychophysical experiments and the results of the application of the selected computational target distincness measures to the target sections of Figures 1 - 6. The rank order induced by the area under the cumulative detection curves in Figure 10 (listed as the R_{Pd} metric in Table III is adopted as the reference visual distinctness order. The rationale for this choice is that this ordering corresponds to a Kolgomorov-Smirnov (K-S) test (see Section 3.1). Rank order permutations of targets that have a comparable visual distinctness (i.e. that are in the same cluster: see Section 3.1) will be called "insignificant". Rank order permutations of elements of different visual distinctness will be called "significant". The rest of this section discusses the psychophysical results (listed in Table III under the header "Perceptual metrics") and the capability of the computational methods (listed in Table III under the header "Digital metrics") to rank order the targets with respect to their visual distinctness.

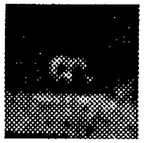
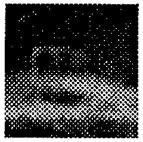
Luminance contrast

The Michelson luminance contrast of the target and its local background are listed in the column with the header C_L in Table III. The results show that the visual distinctness order induced by this metric does not reflect human observer performance in the visual search and detection experiment (see Section 3.1). This conclusion agrees with the finding that observer performance in search and detection tasks is independent of the observer's contrast sensitivity (Task & Pinkus, 1987). This result poses serious problems for models of the human visual search and detection capability that approximate the conspicuity area of a given target by the contrast detection threshold as a function of eccentricity, corresponding to a square or disc with an angular extent equal to that of the target and presented on a homogeneous background (Bowler, 1990; Overington, 1982; Kraiss & Knäeuper, 1982; Waldman et al., 1991). It may explain the fact that these models cannot predict actual observer performance on the acquisition of complex targets in complex backgrounds (e.g. Bijl, 1996; Bijl & Valeton, 1994). Although luminance contrast undoubtedly contributes to the detection probability of a target other factors like the variability of (the structure of and the features in) the local background may dominate target distinctness (e.g. Cole & Jenkins, 1984; Jenkins & Cole, 1982; Nothdurft, 1985b, 1990, 1992, 1993a,b)

Visual lobe

Table III list the visual lobe corresponding to each of the targets in Figures 1 - 6 as a fraction of the total image extent. The results show that this metric attributes the same visual distinctness order to Figures 1 and 2 as the reference metric R_{Pd} , corresponding to the area under the cumulative detection probability curves determined in the search and detection experiment (see section 3.1). However, Figure 2 and Figure 6 are ranked incorrectly. A possible reason for this significant rank order reversal may be the following.

Table III Target distinctness metrics (the upper number in each cell) and induced rank orders (lower number) corresponding to Figs. 1–6. Increasing rank order corresponds to decreasing target distinctness. The rank order induced by the area under the cumulative detection curve (R_{Pd} values, enclosed in circles: from Fig. 10) serves as the reference order. Correct rank orders are displayed in fat boxes, insignificant rank order permutations in thin boxes, and significant rank order permutations are crossed out. (See text for an explanation of the header symbols.)

		Perceptual metrics				**Fit parameters**			**Digital metrics**												
Fig.	**target**	C_L (%)	**lobe** (-)	**ST** (s)	R_{Pd} (%)	τ (s)	t_0 (s)	τ/t_0 (-)	**Doyle** (-)	**Doyle**$_M$ (-)	**nrms** (-)	K_a (%)	K_e (%)	K (%)	**D** (-)	HI$_G$ (-)	HI$_h$ (-)	HI$_v$ (-)	**CAM** (-)	**CTX** (-)	**TVM** (-)	**PD** (-)
1		11.3 ~~5~~	0.17 $\boxed{3/4}$	1.27 $\boxed{1/2}$	94.6 ③	1.64 $\boxed{3}$	0.44 $\boxed{1}$	3.78 $\boxed{3}$	10.9 ~~6~~	9.7 ~~5~~	0.53 $\boxed{3}$	11.2 ~~4~~	3.2 $\boxed{3}$	14.4 ~~4~~	0.43 $\boxed{2}$	0.50 $\boxed{3}$	0.52 $\boxed{2}$	0.39 $\boxed{1}$	1.23 ~~4~~	4.80 $\boxed{2}$	0.61 $\boxed{1}$	5.34 $\boxed{3}$
2		7.8 ~~6~~	0.12 ~~5~~	6.12 $\boxed{4}$	79.3 ④	6.25 $\boxed{4}$	0.73 ~~3~~	8.66 ~~5~~	11.4 ~~5~~	8.4 ~~6~~	0.45 $\boxed{4}$	6.0 ~~6~~	2.6 ~~5~~	8.6 ~~6~~	0.63 $\boxed{4}$	0.57 ~~5~~	0.77 ~~5~~	0.67 ~~6~~	1.96 ~~1~~	2.81 $\boxed{4}$	1.38 ~~5~~	5.15 $\boxed{4}$
3		28.8 $\boxed{3}$	0.21 $\boxed{2}$	1.98 $\boxed{3}$	94.9 ②	1.39 $\boxed{2}$	0.75 ~~4~~	1.85 $\boxed{2}$	20.3 $\boxed{3}$	14.2 $\boxed{3}$	0.60 $\boxed{1}$	9.8 ~~5~~	2.7 ~~4~~	12.5 ~~5~~	0.48 $\boxed{3}$	0.45 $\boxed{1}$	0.69 $\boxed{3}$	0.53 $\boxed{3}$	1.69 $\boxed{3}$	3.13 $\boxed{3}$	1.11 $\boxed{3}$	5.80 $\boxed{2}$
4		25.0 ~~4~~	0.27 $\boxed{1}$	1.27 $\boxed{1/2}$	96.5 ①	0.93 $\boxed{1}$	0.67 $\boxed{2}$	1.61 $\boxed{1}$	13.0 ~~4~~	12.4 ~~4~~	0.57 $\boxed{2}$	19.4 $\boxed{3}$	3.2 $\boxed{2}$	22.6 $\boxed{3}$	0.24 $\boxed{1}$	0.60 ~~6~~	0.51 $\boxed{1}$	0.58 ~~4~~	1.75 $\boxed{2}$	5.34 $\boxed{1}$	0.8 $\boxed{2}$	6.97 $\boxed{1}$
5		36.7 ~~1~~	0.11 $\boxed{6}$	9.15 $\boxed{6}$	66.3 ⑤	9.09 $\boxed{5}$	2.03 $\boxed{6}$	4.47 ~~4~~	22.6 ~~2~~	22.6 ~~2~~	0.33 $\boxed{5}$	30.7 ~~1~~	3.4 ~~1~~	34.1 ~~1~~	0.77 $\boxed{6}$	0.56 ~~4~~	0.70 ~~4~~	0.53 ~~2~~	0.97 $\boxed{6}$	2.51 $\boxed{5}$	1.37 ~~4~~	4.42 $\boxed{5}$
6		36.4 ~~2~~	0.17 ~~3/4~~	8.02 $\boxed{5}$	62.7 ⑥	11.11 $\boxed{6}$	1.12 $\boxed{5}$	10.22 $\boxed{6}$	27.9 ~~1~~	27.7 ~~1~~	0.22 $\boxed{6}$	28.8 ~~2~~	1.6 $\boxed{6}$	30.4 ~~2~~	0.71 $\boxed{5}$	0.47 ~~2~~	0.77 $\boxed{6}$	0.61 $\boxed{5}$	1.11 $\boxed{5}$	1.87 $\boxed{6}$	2.08 $\boxed{6}$	3.51 $\boxed{6}$

A subject can use two different criteria to determine whether the target is visible or not. The first criterium is whether there is anything at the location of the target that has some kind of contrast with the local background. This criterium yields a visual lobe for the *detection* of the target. The second criterium that can be used is whether the spatial structure at the location of the target really originates from the target (can be discriminated as being the target). This criterium yields a visual lobe for the *identification* of the target. In this study the detection lobe was measured. In a follow-up study (Toet et al., 1997) it was found that the recognition lobe induces a visual distinctness rank order that more closely reflects human observer performance in visual search and detection tasks with targets that are not very distinct. Now consider Figure 2 and Figure 6. The target in Figure 6 has a much higher luminance contrast with its local background than the target in Figure 2. However, this target is flanked by trees that also have a high luminance contrast, and that serve as visual distractors (their structure increases the variability of the local background of the target). Moreover, as can be seen from Figure 12, the target in Figure 6 has almost no visible internal structure (it is blob-like), in contrast to the target in Figure 2, which is mainly defined by its edge strength (both external edges from the roof and the front and back sides, and internal edges from the windows).

Summarizing, the visual lobe appears a useful indicator of human performance in a visual search and detection task. The visual distinctness of high contrast targets is probably represented by the detection lobe, whereas the visual distinctness of low contrast targets is probably given by the discrimination lobe.

Mean search time

The mean time the observers need to detect the target in the visual search experiment (described in Section 3.1) is listed in the column with the header ST in Table III. It appears that the visual distinctness order induced by the mean search time yields the correct rank order for the target in Figure 2, and insignificant (within cluster) rank order reversals for all other targets. Summarizing, mean search time appears a reasonable global indicator of human performance in a visual search and detection task.

Empirical model

Search performance is usually expressed as the cumulative detection probability as a function of time, and approximated by

$$P_d(t) = \begin{cases} 0 & : \quad t < t_0 \\ P_\infty (1 - e^{-\frac{t-t_0}{\tau}}) & : \quad t \geq t_0 \end{cases} \tag{41}$$

where

$P_d(t)$ is the fraction correct detections at time t,
P_∞ is the probability of a correct response after an infinite amount of search time,
t_0 is the minimum time required to response, and
τ is a time constant

(e.g. Krendel & Wodinski, 1959; Williams, 1966; Greening, 1976; Akerman & Kinzly, 1979; Ratches et al. , 1981, Rotman et al. , 1989; Waldman et al. , 1991). As Equation 41 clearly shows, search times are not normally distributed. Therefore, simple search time statistics may not correctly characterise observer performance. In most situations search time is restricted and $P_d(t)$ need not approach P_∞ for targets that are hard to find. As a result, the mean of the observed search times may underestimate the true mean (i.e. the mean over the observation period may be less than the mean that would result if the search went on for an infinite amount of time). Other measures may therefore be more suitable to characterise observer search performance, such as the median, geometric mean, and upper and lower quartiles.

In most conditions $P_\infty \approx 1$ so that Equation 41 can be approximated by

$$P_d(t) \;=\; \begin{cases} 0 & : \quad t < t_0 \\ 1 - e^{-\frac{t-t_0}{\tau}} & : \quad t \geq t_0 \end{cases} \tag{42}$$

Williams (1966) defined τ^{-1} as the target conspicuity, and showed this measure can be expressed as the ratio of the search area A (m^2) and the rate C (m^2/s) at which the observer scans the scene, given a certain fixed chance that the observer perceives the target in a single glimpse:

$$\frac{1}{\tau} = \frac{A}{C} \; . \tag{43}$$

Table III lists the parameters τ and t_0 resulting from a least squares fit of Equation 42 to the cumulative detection probability curves shown in Figure 10.

Figure 10 shows that Equation 42 closely describes the experimental data. Table III shows that the inverse of τ indeed strictly decreases with target visual distinctness. This conspicuity metric induces a target distinctness rank ordering that is identical to the order resulting from the R_{Pd} metric.

It has been suggested that t_0 increases with decreasing visual distinctness of the target or with increasing scene complexity (Waldman et al. , 1991). The rationale for this assumption is that an observer tends to need more time to reach a decision when the target is more similar to its background. Table III shows that t_0 may indeed weakly correlate with visual target distinctness. It induces a rank order with four insignificant order reversals, and only 2 significant order reversals.

Table III also shows that the general belief that τ is always much larger than t_0 (the ratio τ/t_0 is always much larger than 1, see e.g. Waldman et al. , 1991), is not true. Therefore, t_0 can not be neglected in Equation 41.

Doyle metric

Table III shows that the Doyle metric (Eq. 6, page 26) and the modified Doyle metric (Eq. 7, with $k = 0.4$) yield similar target distinctness rank orderings. Both show a large number of significant rank order reversals. They both attribute rank orders 1 and 2 to the targets in respectively Figures 6 and 5, which are in fact ranked at the *6th* and *5th* place by the psychophysical R_{Pd} metric. They only yield one insignificant rank order reversal (for the target in Fig. 3).

Summarizing, the results of this pilot experiment suggest that the (modified) Doyle metric is unsuitable to rank order targets in complex natural scenes with respect to their visual distinctness.

Normalised RMS

Table III shows that the $nrms$ metric (Eq. 8, page 27) yields a rank order that is highly similar to the reference rank order based on the psychophysical R_{Pd} metric. The only (insignificant) order reversal is that of the targets in respectively Figs. 3 and 4, that have rank orders 2 and 1 respectively. Inspection of the edge representation of these figures (see Figure 12, page 44) shows that the spatial structure of these targets is indeed highly similar. Both targets are revealed by their roof edges, windows, and the reflections in their head light mirrors. In both cases the background is relatively featureless. As a result both targest have a high feature contrast with their local surround. The cumulative detection probability curves corresponding to these targets (shown in Figure 10, page 22) are very similar and cross each other. A reversal of the rank order of the targets in respectively Figs. 3 and 4 is therefore not significant.

The present results agree with the finding of Kosnik (1995), who showed that search time for airplane silhouettes on complex aireal terrain backgrounds strongly correlates with the $nrms$ metric. Summarizing, the $nrms$ metric appears capable to predict the perceived visual distinctness of targets in complex natural scenes.

Complex contrast

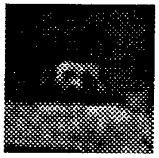
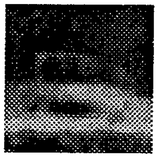
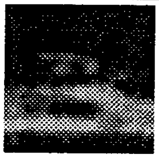
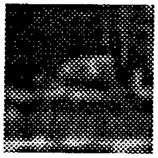
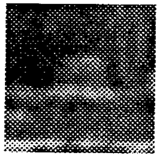
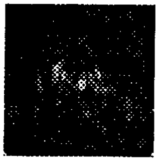
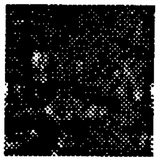
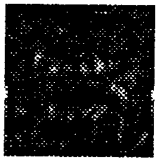
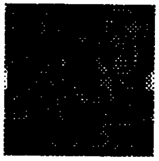
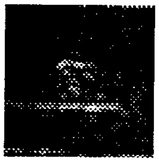
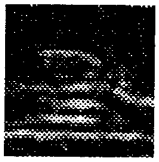
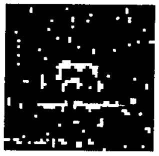
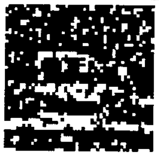
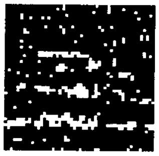
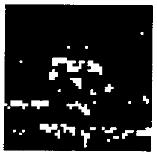
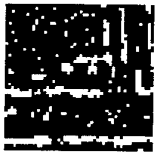
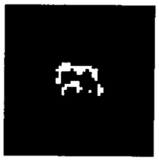
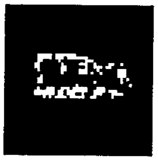
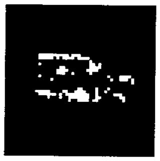
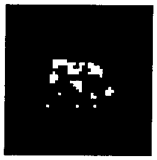
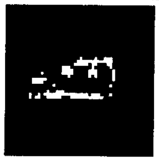
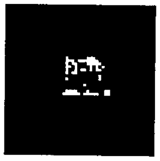
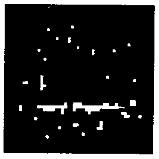
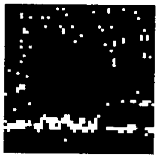
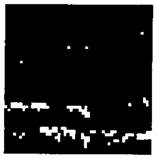
Table III also lists the values of the mean area contrast K_{a} , the mean edge contrast K_{e} , and the resulting complex contrast K (given by Eq. 27, page 33), for the targets in Figures. 1 - 6. The mean area contrast K_{a} performs poorly, which is to be expected, since this measure is similar to the Michelson luminance contrast C_L , which is also unrelated to human visual search performance, as shown above (see page 49). All rank orders computed by this metric are siginificantly out of order relative to the reference order induced by the psychophysical R_{Pd} metric. The mean edge contrast K_{e} produces two correct rank orderings, one insignificant order reversal,and two significant order reversals. The complex contrast measure K resulting from the combination of K_{a} and K_{e} effectively reproduces the incorrect order reversals of its constituting components, and performs worse than either of these measures.

Summarizing, the complex contrast metric K appears not capable to rank order targets in complex natural scenes with respect to their visual distinctness.

Distance metric

Table III shows that the edge distance metric D introduced in this report (and given by Eq. 40, page 46) induces a target distinctness rank order without any significant rank order

Table IV Features in the target and local background sections of Figs. 1–6.

	Fig. 1	**Fig. 2**	**Fig. 3**	**Fig. 4**	**Fig. 5**	**Fig. 6**
target region						
horizontal energy in target region						
vertical energy in target region						
features in target region						
target area of feature image						
background area of feature image						

reversals. Two targets (from Figs. 2 and 4) are ordered correctly. The other targets have been attributed rank orders which do not differ significantly from the reference rank order based on the psychophysical R_{Pd} metric.

Summarizing, the edge distance metric D appears in principle capable to rank order targets in complex natural scenes with respect to their visual distinctness.

Image features and oriented energy

Table IV illustrates the oriented energy approach to the detection of image features. The first row represents the original target sections of Figures 1 - 6. The second and third rows show respectively the horizontal and vertical energy in these images, calculated with Equation 31 (page 36) and using the masks given by Equation 32 (page 36). The fourth row shows the binary feature image $\chi(x, y)$ obtained by taking the pairwise local maximum of corresponding horizontal and vertical energy images (Eq. 33). The last two rows represent the image features in respectively the target support area and the local background. In Figure 4 the local background of the target contains a minimal amount of detail. As a result the target clearly stands out. This may explain the fact that Figure 4 yields the largest visual lobe and the largest area under the cumulative detection curve (fastest overall observer response) in the psychophysical experiments (see Table III). The amount of background detail surrounding the target is larger in Figures 3 and 1. However, the target features are larger and connected in this case, resulting in a Gestalt (form) that induces a perceptual popout effect. This may explain the finding that these two images yield repectively the second and third largest lobe- and R_{Pd} values (see Table III). In Figures 2, 5 and 6 the target is surrounded by a large number of densely packed features. In Figure 2 the target outline (especially the roof) shows linear features that stand out against the random distribution of background features. This image comes on the fourth place in the psychophysical distinctness order induced by the perceptual experiments (see Table III). In Figure 5 the target features are somewhat larger than most of the background features. However, they are not connected and close to some large background features (the tree trunks on the right of the vehicle). In Figure 6 the target features are similar to most of the background features, by which they are closely surrounded. As a result, Figures 5 and 6 end on respectively the fifth and sixth place in the psychophysical rank ordering.

Histogram Intersection

Table III lists the intersection values of respectively

- the graylevel histograms (HI_G),
- the horizontal energy histograms (HI_h), and
- the vertical energy histograms (HI_v)

of the target and its local background (as defined in Table II).

The graylevel histogram intersection yields the correct rank order for Figure 1. For the rest of the targets it yields four significant rank order reversals and one permutation within a cluster of targets of comparable distinctness. The failure of this metric to produce a rank order that correlates with human observer performance can be understood from a careful inspection of Tables II and IV and Figure 16. Figure 16 shows the graylevel histograms of respectively the target section (in red, values increase upwards) and the background region (in green, values increase downwards), and the signed complement of their intersection, computed as the difference between both forementioned histograms (in blue). The blue part of each histogram therefore represents the fraction that is characteristic for the corresponding (target or background) image area (in other words, it indicates the uniqueness of each particular signal). A small entry in the blue histogram at a certain graylevel means that the target and backgrond histograms are very similar (not very distinct) for that graylevel. Overall similarity of the target and background graylevel histograms is reflected in a large amount of blue in these figures, whereas a little amount of blue corresponds to histograms that are very much distinct. Consider the target in Figure 3 which is ranked as the most distinct one (rank order 1) by this approach. Figure 16 shows that the unique fraction of the graylevel histograms corresponding to respectively the target

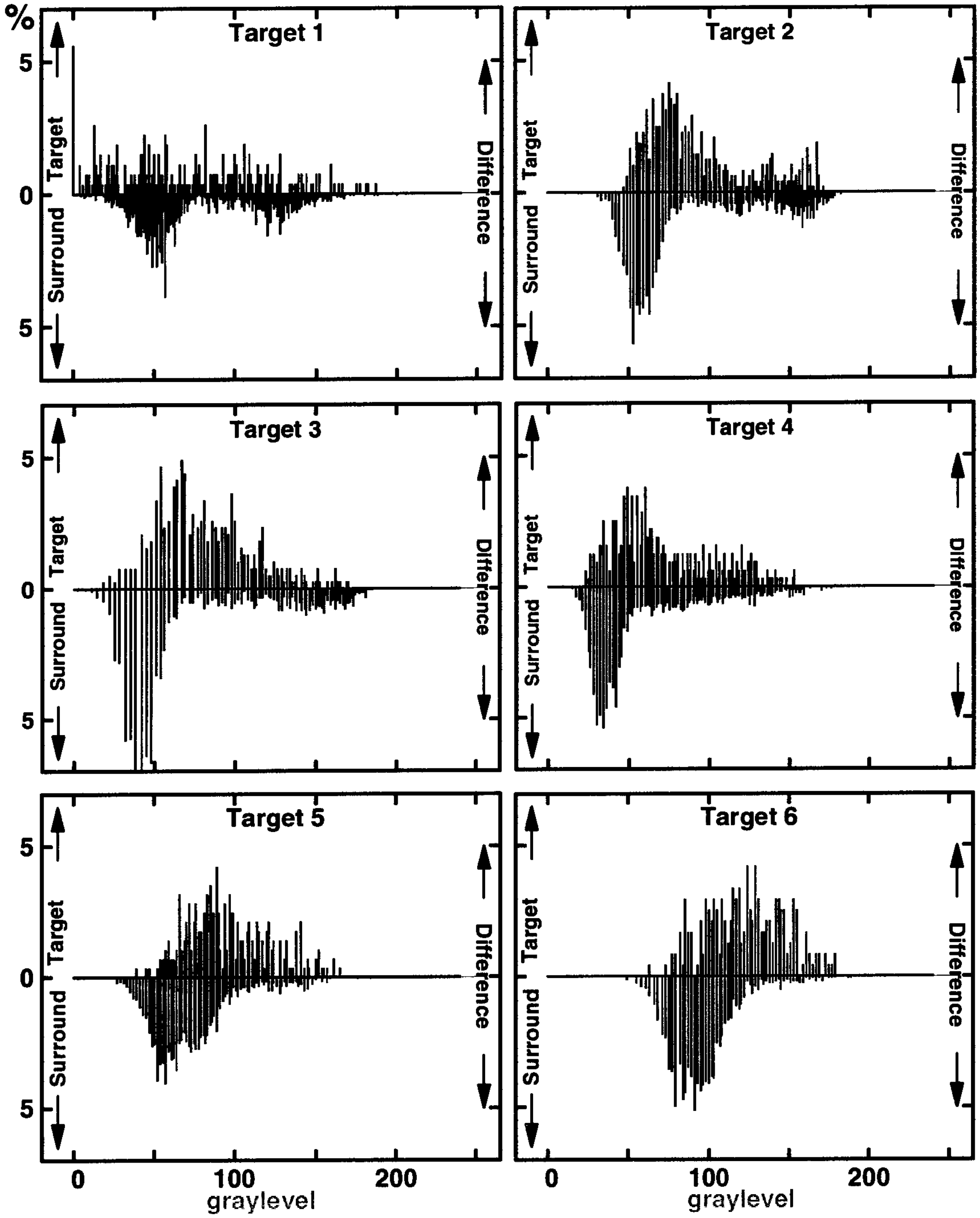


Fig. 16 Graylevel occupation histograms of the target sections of Figs. 1–6.

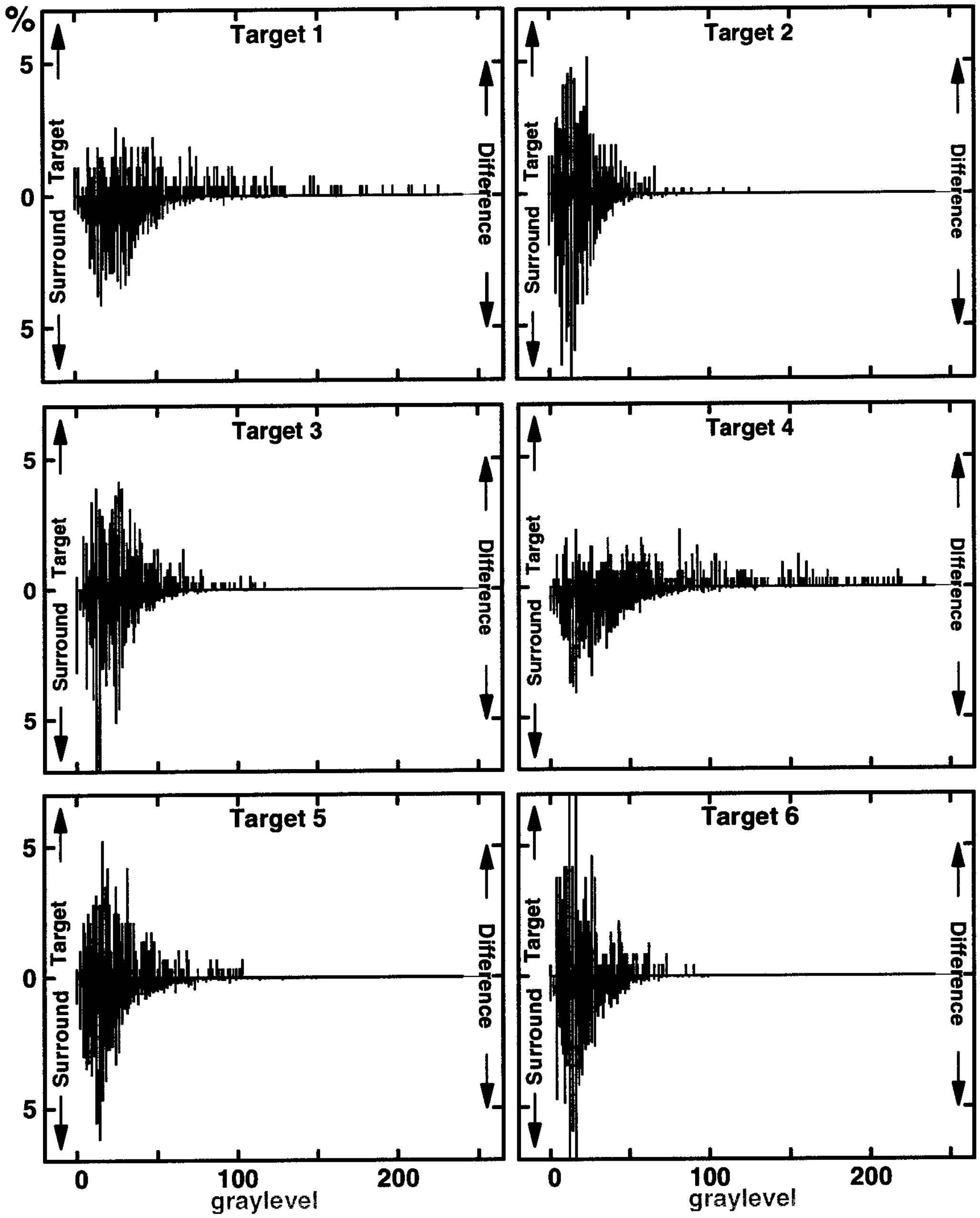


Fig. 17 Horizontal energy histograms of the target sections of Figs. 1–6.

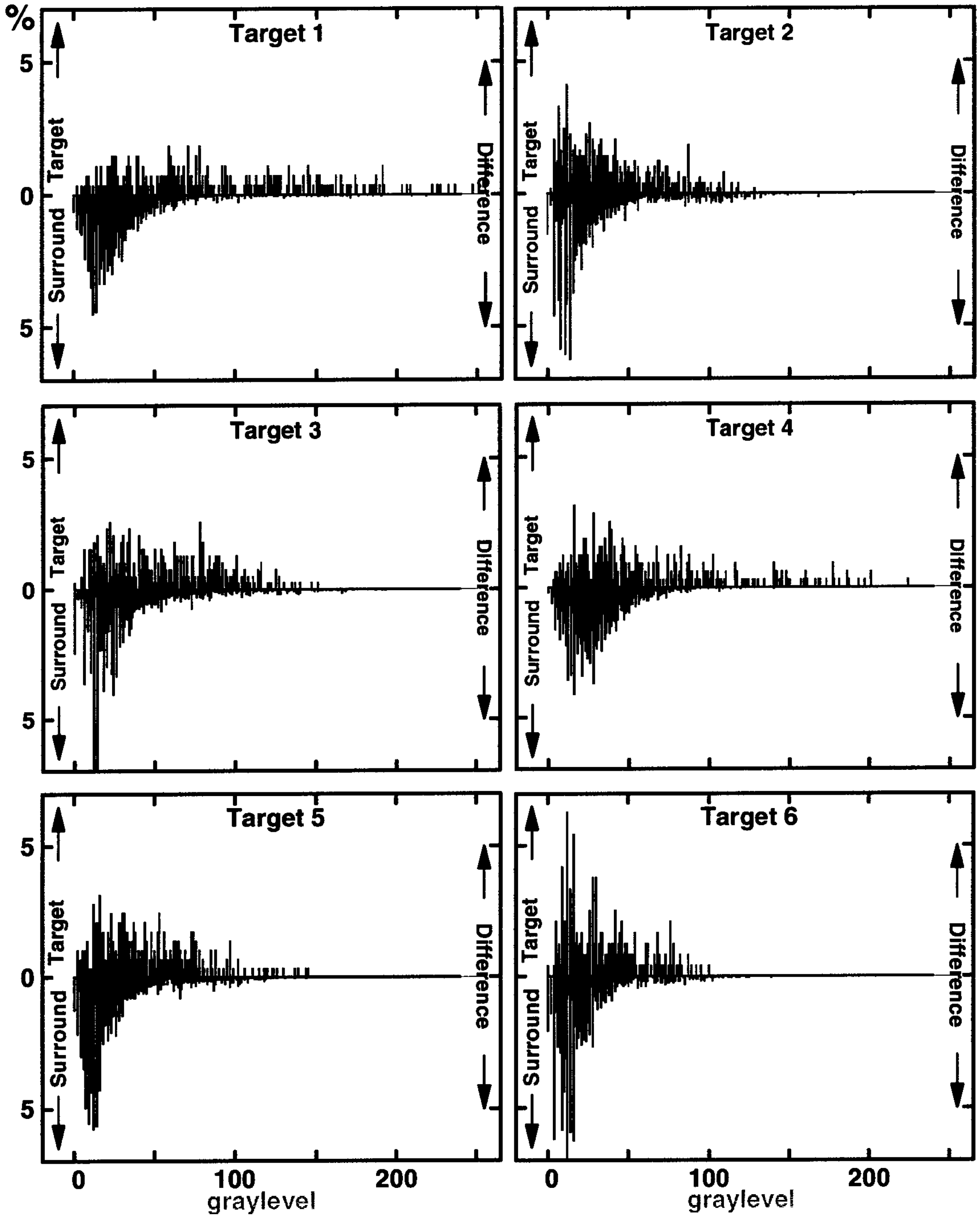


Fig. 18 Vertical energy histograms of the target sections of Figs. 1–6.

and background sections of this figure is indeed large in this case (a large amount of blue is seen in this figure). Also, there is a minimal amount of overlap of the target and background graylevel histograms, with low graylevel values occuring mainly in the background, and high graylevel values occuring mainly in the target support. However, the target in Figure 4, which is actually the most distinct one to the human observers that were tested in this study, is ranked as the least distinct one by the graylevel histogram overlap criterium. As can be seen in Table II, the different graylevels in Figure 4 are highly clustered in blob-like segments. In the background region, low graylevel values occur mainly in the upper part, and high graylevel values occur mainly in the lower part. The reverse holds for the target support, where low graylevel values occur mainly in the lower part, and high graylevel values occur mainly in the upper part. As a result, the (relatively bright) upper part of the target stands out clearly against the (relatively dark) background region. This effect is not captured in the graylevel histogram intersection metric which does include any spatial information. This lack of spatial information in this metric is probably also the reason that the target in Figure 6 is ranked in second order, whereas it is actually the least distinct one to the human observers. Although the target and the background region both have a large unique fraction in their graylevel histograms, there is still a considerable amount of overlap. Inspection of the target section of Figure 6 shows that the target representation is highly blurred in this case. It may therefore be hard to visually distinguish the target from the blurred structures in its surround.

The intersection of the horizontal energy histograms attributes the correct rank order to Figures 4 and 6. For the rest of the targets it yields 2 significant rank order reversals and 2 permutations within a cluster of targets of comparable distinctness. The behaviour of this metric can be understood from an inspection of Tables II and IV and Figure 17. Table II shows that there is not much vertical structure in the background part of the target section. The target itself has a considerable amount of vertical structure. It is therefore a priori likely that a horizontally oriented edge detector (i.e. one that detects vertical edges) is quite capable to segment the target from its local background. The target in Figure 4 is correctly ranked as the most distinct one. Figure 17 shows that this image has indeed the least amount of overlap (the largest unique or blue signal) of the target and background horizontal energy histograms. The target in Figure 6 is correctly ranked as the least distinct one. Figure 17 shows that this image has indeed the largest amount of overlap (the smallest unique or blue signal) of the target and background horizontal energy histograms. Table IV shows that Figures 4 and 6 correspond respectively to the targets with the most detailed and least detailed articulation in the horizontal energy representation. Figure 17 shows that the other targets have a less detailed vertical edge representation, and therefore yield a less pronounced horizontal energy map.

The ordering induced by the intersection of the vertical energy histograms results in 3 significant rank order reversals and 3 (insignificant) permutations within a cluster of targets of comparable distinctness. Vertical energy therefore seems less suited than horizontal energy to rank order the targets with respect to visual distinctness. The behaviour of this metric can be understood from an inspection of Tables II and IV and Figure 18. Table II shows that the target and its local background have about the same amount of horizontally oriented structures (edges). Vertical energy is therefore not suitable to distinguish the target from its surround. Table IV indeed shows similar vertical energy distributions in both regions. The target in Figure 1 is ranked as the most distinct one. This is probably a result of the large horizontal edge strength of the roof of the vehicle. However, the target in Figure 4 also has a large horizontal edge strength, and therefore a strong vertical energy, but is ranked in fourth place. The target in Figure 5 is incorrectly ranked as very distinct (second place), whereas the target in Figure 6 is correctly considered to be indistinct (fifth place). Figure 18 suggests the opposite result, since the roof edge induces an appreciable amount of energy in Figure 5 but not in Figure 6.

Summarizing, the target distinctness rankorder computed with metrics based on histogram intersection correlates poorly with human observer performance. This is probably a result of the fact that this metric contains no (implicit) information about the spatial layout of the scene.

CAMAELEON

As mentioned before (see Section 4.18) CAMAELEON predicts the detection range (km) for the target from an empirically derived function that relates target distinctness to probability of detection. Since the targets in Figures 1 - 6 are at different viewing distances, the target distinctness value is computed here as the ratio of the computed detection range and the actual viewing range (the separation between the camera and the target in the field, as listed in Table I). Although this measure need not correlate with visual distinctness as perceived by the observers in the visual search and detection experiments, it should at least yield a correct rank order.

The target distinctness metric and the resulting rank order computed by the CAMAELEON model are listed in the column with the header CAM in Table III. This model produces a ranking which contains two significant and four insignificant order reversals. It correctly ranks Figures 5 and 6 as the two least distinct targets. However, it suggests that Figure 2 represents the most distinct target, whereas this target is ranked on fourth place based on the outcome of the human observer experiments (see Fig. 10, page 22, and the columns with heading ST and R_{Pd}).

Further experiments with this model (not reported here) indicate that the outcome of the calculations is highly sensitive to the actual definition (size and shape) of both the target and background area. It appears that any conceivable rank order can in principle be obtained by simply adapting the masks for each target. This fact severely degrades the practical value of this model.

Summarizing, although CAMAELEON appears to predict human observer performance to a certain extent, it is highly sensitivity to variations in the definition (size and shape) of the target and background masks. This sensitivity makes the model less useful in practice.

The Cortex Transform

The target distinctness metric and the resulting rank order computed by the Cortex model are listed in the column with the header CTX in Table III. The Cortex model permutes the rank order of Figures 1 and 3, and attributes the correct rank order to all other targets. Since the permuted image pair is in a single cluster of target distinctness, this permutation is not significant.

Summarizing, for the set of test images used in this study, the Cortex model appears to compute a visual target distinctness rank ordering that correlates with human observer performance.

The TARDEC Vision Model

The target distinctness metric and the resulting rank order computed by the TARDEC Vision Model are listed in the column with the header TVM in Table III. The values listed represent the negation of the logarithm of the target detectability metric. As a result, small values correspond to higly distinct targets, and larger values correspond to less distinct targets. TVM correctly ranks Figure 6, which represents the least visible target, last in order. It induces a cyclic permutation on the rank order of Figures 1, 3 and 4. This permutation is not significant, since these targets are in the same cluster of targets with comparable distinctness (their corresponding cumulative detection curves cross each other). The permutation of the rank order of Figures 2 and 5 is significant, because these images belong to different clusters.

Summarizing, the TARDEC Vision Model appears to compute a visual target distinctness metric that predicts (correlates with) human observer performance.

Perceptual Distortion

The target distinctness metric and the resulting rank order computed by the Perceptual Distortion model are listed in the column with the header PD in Table III. This model induces a visual target distinctness rank ordering identical to the one resulting from human observer performance, and therefore shows the best overall performance of all models and metrics tested in this study.

6 GENERAL DISCUSSION

This study is undertaken in an attempt to improve our understanding of which visual target signature characteristics determine search performance in realistic and military relevent complex scenes.

A visual search experiment is performed on a set of complex natural images, each containing a single military vehicle as search target. The area under the resulting cumulative detection probability curves is used to rank order the targets with respect to their visual distinctness. This order is adopted as the reference rank order.

A literature study is performed to select available computational target distinctness measures and early vision models that are a priori most likely to produce results that correlate with human observer performance in visual search and detection tasks. The selected digital image analysis algorithms are applied to quantify visual target distinctness for 6 of the images used in the search experiment. The target visual lobe is also measured on these images.

The visual lobe induces a target distinctness rank order which to some extent agrees with the psychophysical reference rank order. The lack of complete agreement probably results from the fact that the lobe is determined for visual detection, whereas the signatures of targets of low visibility like the ones used in this study are probably better characterised by the lobe for discrimination (recognition). Mean search time also appears a good predictor of overall human visual search performance. The Michelson luminance contrast of the target and its local background do not appear to be strongly related to visual target distinctness.

The cumulative detection curves are closely described by the simple empirical relation given by Equation 42 (see Fig. 10, page 22).

The $nrms$ metric (introduced by Kosnik, 1995) and the distance metric (that is introduced in this report, see Subsection 4.19) show the best overall performance of the first order metrics that are tested in this study. The $nrms$ metric outperfoms the distance metric in its present form. The distance metric can be optimized by the choice of the edge transform, the edge threshold level, and the distance transform. However, the $nrms$ metric is much simpeler to compute, and not very sensitive to the actual choice of the target and background support, which makes it very useful for practical applications..

The Cortex (Section 4.15) TARDEC (Section 4.14), and Perceptual Distortion (Section 4.16) early vision models all compute visual target distinctness metrics that appear to correlate with human observer performance. For the present set of test images the Perceptual Distortion model induces a visual target distinctness rank ordering identical to the one resulting from human observer performance, and therefore shows the best overall performance of all models and metrics tested in this study.

7 CONCLUSIONS

The following conclusions can be drawn from the experimental paradigms and the (restricted) set of test images used in this pilot study.

1. The visual lobe appears a useful indicator of human performance in a visual search and detection task.
2. Complex contrast, the (modified) Doyle metric, and metrics based on histogram intersection correlate poorly with human observer performance.
3. The CAMAELEON model is highly sensitivity to variations in the definition (size and shape) of the target and background masks. This sensitivity makes the model less useful in practice.
4. Models of the early human visual system (Cortex, TARDEC, Perceptual Distortion), the $nrms$ metric, and the edge distance metric D all seem to induce a visual distinctness rank ordering that agrees with human visual perception.
5. The Perceptual Distortion model induces a visual target distinctness rank ordering identical to the one resulting from human observer performance, and therefore shows the best overall performance of all models and metrics tested in this study.

8 FURTHER RESEARCH

A follow-up study is planned to further validate the computational target distinctness metrics that perform best in the present pilot study. The new study will include a larger number of images. The results of the computational measures will be compared with the results from a psychophysical search and detection experiment through a paired rank order test.

Acknowledgements

The work reported here was performed in close collaboration with the Armstrong Laboratory Optical Radiation Division (AL/OEO) at Brooks AFB, TX. The author wishes to acknowledge the stimulating discussions with Dr. LtC Frank Cheney, Dr. Bill Kosnik and the other members of AL/OEO.

The author would like to thank the following persons for their active help with the model calculations:
Al Ahumada (NASA Ames Research Center, Moffett Field, CA),
Grant Gerhart and Tom Meitzler (US Army Tank-automotive and Armaments Command Research, Development and Engineering Center, Warren, MI),
Bill Kosnik (TASC, AL/OEOP, Brooks AFB, San Antonio, TX, USA),
Javier Martínez-Baena, Joaquin Fdez-Valdivia, and Jose A. Garcia, (Dpto. Ciencias de la Computacion e Inteligencia Artificial, Escuela Tecnica Superior de Ingenieria Informatica Universidad de Granada, Granada, Spain).

REFERENCES

Adelson, E.H. and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A, 2,* 284–299.

Ahumada, A.J. and Beard, B.L. (1996). Object detection in a noisy scene. In Rogowitz, B. and Allebach, J. (Eds.), *SPIE Proceedings on Human Vision and Digital Display VII, SPIE Vol. 2657,* (paper 23). Bellingham, WA.: SPIE. Available via World Wide Web from `http//vision.arc.nasa.gov`.

Ahumada, A.J., Rohaly, A.M., and Watson, A.B. (1995). Image discrimination models predict object detection in natural backgrounds. Supplement to Investigative Ophthalmology and Visual Science, 36(4), abstract 2013, p. S439. Available via World Wide Web from `http//vision.arc.nasa.gov`.

Akerman, A. and Kinzly, R.E. (1979). Predicting aircraft detectability. *Human Factors, 21(3),* 277–291.

Alkhateeb, W.F., Morris, R.J. and Ruddock, K.H. (1990a). Effects of complexity on simple spatial discriminations. *Spatial Vision, 5,* 129–141.

Ballard, D.H. and Brown, C.M. (1992). Principles of animate vision. *Computer Vision, Graphics and Image Processing: Image Understanding, 56(1),* 3–21.

Baraldi, A. and Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level coocurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing 33(2),* 293–304.

Barten, P.G.J. (1993). Spatiotemporal model for the contrast sensitivity of the human eye and its temporal aspects. In Rogowitz, B. and Allebach, J. (Eds.), *SPIE Proceedings on Human Vision, Visual Processing and Digital Display IV, SPIE Vol. 1913* (pp. 2–14). Bellingham, WA.: SPIE.

Beck, J. (1966a). Perceptual grouping produced by changes in orientation and shape. *Science, 154,* 538–540.

Beck, J. (1966b). Effects of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300–302.

Beck, J. (1972). Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology, 85,* 1–19.

Beck, J. (1982). Textural segmentation. In Beck, J. (Ed.), *Organization and representation in perception* (pp. 285–317). Hillsdale, NJ: Erlbaum.

Bergen, J.R. and Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 696–698.

Bijl, P. and Valeton, J.M. (1994). *Evaluation of target acquisition model "TARGAC" using "BEST TWO" observer performance data* (Report TNO-TM 1994 A22). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Bijl, P. (1996). *Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictionsfor air-to-surfcae FLIR imagery* (Report TM-96-A037). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Birkmire, D.P., Karsh, R., Barnette, B.D., and Pillalamarri, R. (1992). Target acquisition in cluttered environments. In *Proceedings of the 36th Meeting of the Human Factors Society* (pp. 1425–1429).

Bloomfield, J.R. (1972). Visual search in complex fields: size differences between target disc and surrounding discs. *Human Factors, 14,* 139–148.

Borgefors, G. (1986) Distance transformations in digital images. *Computer Vision, Graphics and Image Processing, 34,* 344–371.

Bowler, Y.M. (1990). Towards a simplified model of visual search. In Brogan, D. (Ed.), *Visual Search* (pp.303–309). London, UK: Taylor & Francis.

Braje, W.L., Tjan, B.S., and Legge, G.E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research, 35(21),* 2955–2966.

Burr, D.C. and Morrone, M.C. (1990). Feature detection in biological and artificial visual systems. In C. Blakemore (Ed.), *Vision: coding and efficiency* (pp. 185–194). Cambridge, MA: Cambridge University Press.

Burt, P. (1988a). Algorithms and architecture for smart sensing. In *Image Understanding Workshop.* Cambridge, MA: DARPA.

Burt, P. (1988b). Smart sensing with a pyramid vision machine. *Proceedings IEEE, 76,* 1006–1015.

Caelli, T. and Moraglia, G. (1986). On the detection of signals embedded in natural scenes. *Perception & Psychophysics, 39(2),* 87–95.

Campbell, F.W., Kulikowski, J.J., and Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *Journal of Physiology, 187,* 427–436.

Carlson, C.R. and Cohen, R.W. (1980). A simple psychophysical model for predicting the visibility of displayed information. *Proceedings Society for Information Display,* 21(3), pp. 229–246.

Cathcart, J.M., Doll, T.J. and Schmieder, D.E. (1989). Target detection in urban clutter. *IEEE Transactions on Systems, Man and Cybernetics SMC, 19,* 1242–1250.

Cole, B.L. and Jenkins, S.E. (1984). The effect of variability of background elements on the conspicuity of objects. *Vision Research, 24,* 261-270.

Copeland, A.C., Trivedi, M.M., and McManamey, J.R. (1996). Evaluation of image metrics for target discrimination using psychophysical experiments. *Optical Engineering, 35(6),* 1714–1722.

Daniellson, P.E. (1980). Euclidian distance mapping. *Computer Vision, Graphics and Image Processing, 14,* 227–248.

Dick, M., Ullman, S. and Sagi, D. (1987). Parallel and serial processes in motion detection. *Science, 237,* 400-402.

Doll, T.J., McWhorter, S.W., and Schmieder, D.E. (1993). Target and background characterization based on a simulation of human pattern perception. In *Proceedings SPIE Conference on Characterization, Propagation, and Simulation of Sources and Backgrounds III, SPIE Vol. 1967* (pp. 432–454). Bellingham, WA: SPIE.

Doll, T.J., McWhorter, S.W., and Schmieder, D.E., and Wasilewski, A.A. (1995). Simulation of selective attention and training effects in visual search and detection. In E. Peli (Ed.), *Vision models for target detection and recognition* (pp. 396–418). Singapore: World Scientific.

Engel, F.L. (1971). Visual conspicuity. Directed attention and retinal locus. *Vision Research, 11,* 563–575.

Engel, F.L. (1974). Visual conspicuity and selective background interference in eccentric vision. *Vision Research, 14,* 459–471.

Engel, F.L. (1977). Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research, 17,* 95–100.

Ennesser, F. and Medioni, G. (1995). Finding Waldo, or focus of attention using local color information. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 17(8),* 805–809.

Fdez-Vidal, X.R., Garcia, J.A. and Fdez-Valdivia, J. (1996a). Using models of feature perception in distortion measure guidance (Report DECSAI 96-03-33). Granada, Spain: Computer Science Department, University of Granada. Available via ftp at: `ftp://decsai.ugr.es/pub/diata/tech_rep/TR960333.ps.Z`

Fdez-Vidal, X.R., Garcia, J.A. and Fdez-Valdivia, J. (1996b). The role of integral features for perceiving image distortion (Report DECSAI 96-03-39). Available via ftp at: `ftp://decsai.ugr.es/pub/diata/tech_rep/TR960339.ps.Z.`

Foster, D.H. and Ward, P.A. (1991). Asymmetries in oriented line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London B, 243,* 75–81.

Gerhart, G., Meitzler, T., Sohn, E., Witus, G., Lindquist, G., and Freeling, J.R. (1995). Early vision model for target detection. In *Proceedings SPIE Conference on Infrared Imaging Systems: Design, Analysis, Modeling, and Testing VI, SPIE Vol. 2470* (pp. 12–23). Bellingham, WA: SPIE.

Golomb, B., Andersen, R.A., Nakayama, K., MacLeod, D.I.A., and Wong, A. (1985). Visual thresholds for shearing motion in monkey and man. *Vision Research, 25,* 813–820.

Graham, N. (1991). Complex channels, early nonlinearities, and normalization in texture segregation. In Landy, M.S. and Movshon, J.A. (Eds.), *Computational models of visual processing* (pp. 273–290). Cambridge, MA: MIT Press.

Graham, N., Beck, J., and Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: effects of sign and amount of contrast. *Vision Research, 32,* 719–743.

Greening, C.P. (1976). Mathematical modeling of air-to-ground target acquisition. *Human Factors, 18(2),* 111–148.

Greenlee, M.W. and Magnussen, S. (1988). Interactions among spatial frequency and orientation channels adapted concurrently. *Vision Research, 28,* 1303–1310.

Haralick, R.M., Shanmugam, K., and Dinstein, I. (1976). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics SMC, 3,* 610–621.

Hecker, R. (1992). Camaeleon - Camouflage assessment by evaluation of local energy, spatial frequency and orientation. In *Proceedings SPIE Conference on Characterization, Propagation, and Simulation of Sources and Backgrounds II, SPIE Vol. 1687* (pp. 342–349). Bellingham, WA: SPIE.

Heckmann, T, Witus, G., Meitzler, T., Gerhart, G. and Sohn, E. (1995). *Progress toward an automotive conspicuity enhancement invention tool (ACEIT)* (Report CO-29). Warren, MI: GM Research and Development Center.

Heeger, D.J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America, A, 4(8),* 1455–1471.

Heeger, D.J. (1988). Optical flow using spatiotemporal filters. *International Journal of Computer Vision, 1,* 279–302.

Jenkins, S.E. and Cole, B.L. (1982). The effect of the density of background elements on the conspicuity of objects. *Vision Research, 22,* 1241-1252.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell Systems Technical Journal, 39,* 1125–1162.

Julesz, B. (1964) Binocular depth perception without familiarity cues. *Science, 145,* 356–362.

Julesz, B. (1971). *Foundations of cyclopean perception.* Chicago, Ill: Chicago University Press.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232,* 34–43.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience, 7,* 41–45.

Kastner, S., Nothdurft, H.-C., and Pigarve, I.N. (1997). Neural correlates of pop-out in cat striate cortex. *Vision Research, 37,* 371–376.

Kelly, D.H. (1972). Adaptation effects on spatio-temporal sine-wave thresholds. *Vision Research, 12,* 89–101.

Koch, C. and Ullman, S. (1985). Shifts in visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kooi, F.L. and Valeton, J.M. (1997). *Quantifying the conspicuity of objects in real scenes* (Report TNO-TM 1997, in preparation). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Kosnik, B. (1995). Quantifying target contrast in target acquisition research. In Peli, E. (Ed.), *Vision models for target detection and recognition* (pp. 380–395). Singapore: World Scientific.

Kovesi, P. (1995). *Image features from phase congruency* (Report 95/4). Nedlands, Australia: Department of Computer Science, The University of Western Australia. Available via ftp at cs.uwa.edu.au at /pub/techreports/95/4.ps.gz.

Kraiss, K.-F. and Knäeuper, A. (1982). Using visual lobe area measurements to predict visual search performance. *Human Factors, 24,* 673–682.

Krendel, E.S. and Wodinsky, J. (1960). Visual search in an unstructured visual field. *Journal of the Optical Society of America, 50,* 562–568.

Lakshmanan, S., Meitzler, T., Sohn, E., and Gerhart, G. (1995). Simulation and comparison of infra-red sensors for automotive collision avoidance. In *Proceedings IHVS and Advanced Transportation Systems (SP-1076)* (pp. 99–104). Warrendale, PA: Society of Automotive Engineers.

Lamdan, Y., Schwartz, J.T., and Wolfson, H. (1988). On recognition of 3-d objects from 2-d images. In *Proceedings of the IEEE International Conference on Robotics and Automation, Philadelphia* (pp. 1407–1413).

Lillesæter, O. (1993). Complex contrast, a definition for structured targets and backgrounds. *Journal of the Optical Society of America, A, 10,* 2453–2457.

Lowitz, G.E. (1983). Can a local histogram really map texture information? *Pattern Recognition, 16(2),* 141–147.

Lowitz, G.E. (1984). Mapping the local information content of a spatial image. *Pattern Recognition, 17(5),* 545–550.

Marr, D. (1982). *Vision.* San Francisco, USA: Freeman.

Martinez-Baena, J., Fdez-Valdivia, J., and Garcia, J.A. (1996). *Distortion measures based on a data-driven multisensor organization* (Report DECSAI 96-03-17). Granada, Spain: Computer Science Department, University of Granada. Available via ftp at decsai.ugr.es/pub/diata/tech_rep/TR960317.ps.Z.

Meitzler, T., Gerhart, G., Sohn, E., Witus, G., Heckmann, T, Cusumano, E., Polewarcyk, J. (1995). *Adapting an army vision model for measuring armored-vehicle camouflage to evaluating the conspicuity of civilian vehicles* (Report). Warren, MI: US Army Tank-automotive and Armaments Command, Research, Development and Engineering Center.

Mel, B.W. (1996). *SEEMORE: combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition.* Available via ftp at quake.usc.edu/pub/mel/papers/mel.seemore.ps.gz.

Milanese, R. (1993). *Detecting salient regions in an image: from biology to implementation.* (Ph.D. Thesis). Genova, Switzerland: Computer Science Department, University of Genova. Available via ftp at cui.unige.ch as pub/milanese/thesis.

Milanese, R., Bost, J-M. and Pun, T. (1992). A bottom-up attention system for active vision. In

Proceedings of the 10th European Conference on Artificial Intelligence (pp. 808–810). New York: John Wiley.

Milanese, R., Wechsler, H., Gil, S., Bost, J-M. and Pun, T. (1994). Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 781-785).

Mitchell, D.E. and Wilkinson, F. (1974). The effect of early astigmatism on the visual resolution of gratings. *Journal of Physiology, 243,* 739–756.

Moravec, H.P. (1977). Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence* (pp. 584–590). Cambridge, MA.

Morrone, M.C. and Burr, D.C. (1988). Feature detection in human vision: a phase dependent energy model. *Proceedings of the Royal Society of London B, 235,* 221–245.

Morrone, M.C. and Owens, R. (1987). Feature detection from local energy. Pattern Recognition Letters, 6, 303–313.

Morrone, M.C., Ross, J., Burr, D.C., and Owens, R. (1986). Mach bands are phase dependent. *Nature, 324,* 250–253.

Nakayama, K., Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264–265.

Nakayama, K., Silverman, G.H., MacLeod, D.I.A. and Mulligan, J. (1985). Sensitivity to shearing and compressive motion in random dots. *Perception, 14,* 225–238.

Nakayama, K. and Tyler, C.W. (1981). Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research, 21,* 427–433.

Nothdurft, H.C. (1985b). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research, 25,* 1957–1968.

Nothdurft, H.C. (1987). Perceptive fields of direction selective neurons in the human visual system. In N. Elsner and O.D. Creutzfeldt (Eds.), *New frontiers in brain research* (p. 180). Stuttgart, GE: Thieme Verlag.

Nothdurft, H.C. (1990). Texton segregation by associated differences in global and local luminance distribution. *Proceedings of the Royal Society of London B, 239,* 295–320.

Nothdurft, H.C. (1991b). Texture segregation and pop-out from orientation contrast. *Vision Research, 31,* 1073–1078.

Nothdurft, H.C. (1991c). Texture segmentation and pop-out from orientation contrast. *Vision Research, 31,* 1073–1078.

Nothdurft, H.C. (1992). Feature analysis and the role of similarity in pre-attentive vision. *Perception & Psychophysics, 52,* 355–375.

Nothdurft, H.C. (1993a). Saliency effects across dimensions in visual search. *Vision Research, 33,* 839–844.

Nothdurft, H.C. (1993b). The role of features in preattentive vision: comparison of orientation, motion and color cues. *Vision Research, 33,* 1937–1958.

Nothdurft, H.C. (1993c). The conspicuousness of orientation and motion contrast. *Spatial Vision, 7,* 341-363.

Olshausen, B.A., Anderson, C.H. and Essen, D.C. van (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13(11),* 4700–4719.

Olson, R.K. and Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology, 83,* 1-21.

Overington, I. (1982). Towards a complete model of photopic visual threshold performance. *Optical Engineering, 21,* 2–13.

Owens, R. (1994). Feature-free images. Pattern Recognition Letters, 15, 35–44.

Paglieroni, D.W. (1992). Distance transforms: properties and machine vision applications. *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing, 54,* 56–74.

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America, A, 7,* 2032–2040.

Pratt, W.K. (1991). *Digital image processing, 2nd ed.* New York: Wiley.

Quick, R.F. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16,* 65–67.

Ratches, J.A., Lawson, W.R., Shields, F.J., Hoover, C.W., Obert, L.P., Rodak. S.P., and Sola, M.C. (1981). *Status of Sensor Performance Modelling at NV&EOL.* Fort Belvoir, VA: Night Vision & Electro-Optics Laboratory.

Reisfeld, D., Wolfson, H., and Yeshurun, Y. (1990). Detection of interest points using symmetry. *Proceedings of the Third International Conference on Computer Vision* (pp. 62–65). Washington, USA: IEEE Computer Society Press.

Reisfeld, D., Wolfson, H., and Yeshurun, Y. (1995). Context-free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision, 14,* 119–130.

Robbins, B. and Owens, R. (1994). *The 2D local energy model* (Report 94/5). Nedlands, Australia: Department of Computer Science, The University of Western Australia.

Ronse, C. (1995). *The phase congruence model for edge detection in two-dimensional pictures: a mathematical study* (Report 95/11). Strasbourg: LSIIT, Universite Louis Pasteur. Also available via World Wide Web from `http://dpt-info.u-strasbg.fr/~ronse/` as `file://dpt-info.u-strasbg.fr/pub/recherche/Vision/rap95-11.ps.Z.`

Rosenfeld, A. and Kak, A.C. (1982). *Digital Picture Processing, 2nd. ed., Vol.1 and 2.* New York: Academic Press.

Rotman, S.R., Gordon, E.S., and Kowalczyk, M.L. (1989). Modeling human search and target acquisition performance: I. First detection probability in a realistic multitarget scenario. *Optical Engineering, 28(11),* 1216–1222.

Rotman, S.R., Kowalczyk, M.L. and George, V. (1994a). Modeling human visual search and target acquisition performance: fixation-point analysis. *Optical Engineering, 33(11),* 3803–3809.

Rotman, S.R., Tidhar, G., and Kowalczyk, M.L. (1994b). Clutter metrics for target detection systems. *IEEE Transactions on Aerospace Electronic Systems, 30,* 81–91.

Rybak, I.A., Golovan, A.V., and Gusakova, V.I. (1993). Behavioral model of visual perception and recognition. In *Proceedings of the SPIE Conference on Human Vision, Visual Processing, and Digital Display IV, SPIE Vol. 1913* (pp. 548–560). Bellingham, WA: SPIE.

Sachtler, W.L. and Zaidi, Q. (1995). Visual processing of motion boundaries. *Vision Research, 35,* 807–826.

Sagi, D. and Julesz, B. (1987). Short-range limitations on detection of feature differences. *Spatial Vision, 2,* 39–49.

Schmieder, D.E. and Weathersby, M.R. (1983). Detection performance in clutter with variable resolution. *IEEE Transactions on Aerospace and Electronic Systems, 19(4),* 622–630.

Serra, J. (1982). *Image analysis and mathematical morphology.* New York, USA: Academic Press.

Shirvaikar, M.V. and Trivedi, M.M. (1992). Developing texture-based image clutter measures for object detection. *Optical Engineering, 31,* 2628–2639.

Skjervold, J. (1995). Extensions of the US Night Vision Laboratory model for thermal viewing systems on structural targets and backgrounds in cluttered scenes. In *Proceedings SPIE Conference on Targets and Backgrounds: characterization and representation, SPIE Vol. 2469* (pp. 568–575). Bellingham, WA: SPIE.

Swain, M.J.and Ballard, D.H. (1991). Color indexing. *International Journal of Computer Vision,* *7(1)*, 11-32.

Task, H.L. and Pinkus, A.R. (1987). Contrast sensitivity and target recognition performance: a lack of correlation. In *Proceedings of the SID International Symposium 1987* (127-129). New York, USA: Palisades Institute for Research Services Inc.

Tidhar, G., Reiter, G., Avital, Z., Hadar, Y., Rotman, S.R., George,V., and Kowalczyk, M.L. (1994). Modeling human search and target acquisition performance: IV. detection probability in the cluttered environment. *Optical Engineering, 33*, 801-808.

Toet, A., Bijl, P., F.L. Kooi, F.L., and J.M. Valeton, J.M. (1997). *Image data set for testing search and detection models* (Report TNO-TM 1997, in preparation). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Treisman, A. and Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95*, 15-48.

Trivedi, M.M., Harlow, C.A., Conners, R.W., and Goh, S. (1984). Object detection based on gray level coocurrence. *Computer Vision, Graphics and Image Processing, 28*, 199-219.

Tsotsos, J.K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences, 13*, 423-469.

Tsotsos, J.K. (1993). An inhibitory beam for attentional selection. In L. Harris, and M. Jenkin (Eds.), *Spatial vision in humans and robots* (pp. 313-331). Cambridge, MA: Cambridge University Press.

Tsotsos, J.K. (1994). Towards a computational model of visual attention. In T.V. Papathomas, C. Chubb, A. Gorea, and E. Kowler (Eds.), *Early Vision and Beyond* (pp. 207-218). Cambridge, MA: MIT Press/Bradford Books.

Venkatesh, S. and Owens, R. (1990). On the classification of images features. *Pattern Recognition Letters, 11*, 339-349.

Verwer, B.J.H. (1991) Local distances for distance transformations in two and three dimensions. *Pattern Recognition Letters, 12*, 671-682.

Waldman, G., Wootton, J. and Hobson, G. (1991). Visual detection with search: an empirical model. *IEEE Transactions on Systems, Man and Cybernetics SMC, 21(3)*, 596-606.

Waldman, G., Wootton, J., Hobson, G. and Luetkemeyer, K. (1988). A normalized clutter measure for images. *Computer Vision, Graphics and Image Processing, 42*, 137-156.

Watson, A.B. (1983). Detection and recognition of simple spatial forms. In O.J. Braddick and A.C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 100-114). Berlin, GE: Springer.

Watson, A.B. (1987). The Cortex transform: rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing, 39*, 311-327.

Watson,A.B. and Ahumada, A.J. (1985). Model of human visual motion sensing. *Journal of the Optical Society of America, A, 2(2)*, 322-341.

Wertheim, A.H. (1989). *A quantitative conspicuity index; theoretical foundation and experimental validation of a measurement procedure* (Report C-20, in Dutch). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Williams, L.G. (1966). Target conspicuity and visual search. *Human Factors, 8*, 80-92.

Wilson, H.R. and Richards, W.A. (1992). Curvature and separation discrimination at texture boundaries. *Journal of the Optical Society of America, A, 9*, 1653-1662.

Wilson, H.R., McFarlane, D.K., and Phillips, G.C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23*, 873-882.

Witus, G., Cohen, M., Cook, T., Elliott, M., Freeling, J.R., Gottschalk, P., Lindquist, G. (1995a). *TARDEC visual model version 2.1.1 analyst's manual* (Report OMI-552). Ann Arbor, MI:

OptiMetrics Inc.

Witus, G., Heckmann, T.R., Meitzler, T., Gerhart, G., and Sohn, E. (1995b). *Evaluating an army camouflaged vehicle visual signature model for measuring civilian vehicle conspicuity* (Report GM R&D-8350). Warren, MI: General Motors, Research and Development Center.

Witus, G., Heckmann, T.R., Meitzler, T., Gerhart, G., and Sohn, E. (1995c). Evaluating a computer model of search and detection in complex scenes. *Supplement to Investigative Ophthalmology and Visual Science, 36(4),* abstract # 4128.

Wolfe, J.M. (1992). The parallel guidance of visual attention. *Current Directions in Psychological Science, 1,* 125–128.

Wolfe, J.M. (1994a). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J.M. (1994b). Visual search in continuous naturalistic stimuli. *Vision Research, 34,* 1187–1195.

Wolfe, J.M. and Cave, K.R. (1989). Deploying visual attention: The guided search model. In T. Troscianko and A. Blake (Eds.), *AI and the eye* (pp. ??). London, UK: Wiley and Sons.

Wolfe, J.M., Cave, K.R. and Franzel, S.L. (1989). Guided Search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Yeshurun, Y. and Schwartz, E.L. (1989). Shape description with a space-variant sensor: algorithm for scan-path, fusion, and convergence over multiple scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 11(11),* 1217–1222.

Zamperoni, P. (1995). Model-free texture segmentation based on distances between first-order statistics. *Digital Signal Processing, 5,* 197–225.